Institute of Educational Sciences
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES
TECHNICAL EXPERT PANEL REPORT

REMOTE SENSING TO ESTIMATE US K-12 PHYSICAL PLANT

Table of Contents

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

REMOTE SENSING TO ESTIMATE U.S. K–12 PHYSICAL PLANT

EXECUTIVE SUMMARY

While the National Center for Education Statistics (NCES) collects annually and maintains administrative data on all K-12 schools in the United States, there is no comparable collection of information on schools' physical plant (buildings, grounds and other infrastructure necessary for each school). Yet decision-makers at all levels need this kind of information as they set policy and develop facility plans. Remote sensing imagery could provide an avenue to providing information on a national basis, identifying school buildings and estimating the usable space for teaching/learning and other school activities. NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of technical experts to consider how to respond to the need to estimate the physical plant for US K-12 schools, what role remote sensing might play, and what resources would have to be located or created to use these data efficiently.

The panel met via teleconferences with an in-person meeting at NCES on 19-20 February 2020.

The panel separated buildings used in instruction from exterior grounds, athletic facilities and parking lots for initial consideration. Staging the data base development and incorporating existing (federal, if possible) data bases would allow NCES to benefit from raw data, image data processing and calculations already developed for other purposes. LIDAR (LIght Detection And Ranging) would provide 3-dimensional rather than 2-dimensional remote sensing images. However, LIDAR research, implementation and applications are far less advanced for purposes similar to the estimation of K-12 school facilities. Therefore, the panel focused on determining the feasibility of a plan that would rely on remote sensing imagery coupled with administrative, observational and other alternative data sources for the creation and maintenance via annual updating of a record of total 2-dimensional space ("footprint") and total square footage "under roof."

A preliminary template for a data record was described that would inventory individual structures and also summarize information for each school or campus in the case of multiple schools on shared or adjoining space. A multiple stage approach was outlined for the acquisition of the remote sensing data, image creation and calculations. Requirements for a systematic updating process were sketched. The panel recognized that the diversity and level of technical skill for this project would require specialized expertise not available within NCES.

Recommendations

1. Estimation of US K-12 physical plant to be undertaken in stages.

Stage 1 to be limited to comprehensive data for a pilot set of states with both administrative information on buildings for each school (footprint, square footage under roof) and parcel data in a usable form to be integrated with remote sensing imagery. Stage 1 to focus on definition of a data base (and associate primary data base) for two-dimensional information (footprints), and to include validation, development of

diagnostics and adjustments, with a view to detection of multi-story buildings, verification and consequent adjustments.

Stage 2 to scale up to national level with incorporation of updating processes.

Stages 3 and 4 to be reconsidered with respect to efficient estimation for multi-story buildings and to available options for detection and inventory of non-structure facilities including athletic fields, playgrounds and parking lots.

2. Maximized use of existing data bases and existing software.

In particular, the federal data base (USA Structures – being developed at Oak Ridge National Laboratory for FEMA), and data bases with parcel boundaries geo-inscribed can avoid redevelopment of software and instability of data sources over time.

3. Annual updating including auxiliary information.

A formal system for updating on an annual basis to be built, independent of remote sensing, to reflect changes in school facilities between revisions to remote sensing imagery data.

PREFACE

The National Center for Education Statistics (NCES) maintains a mandated data base that is a complete inventory of educational institutions in the United States. Information in this data base includes a broad collection of administrative information. It includes GIS point location for facilities but does not include specific information on the physical plant such as total square footage of usable permanent space, number of temporary buildings, or exterior space information such as athletic fields or playgrounds. Neither is this information routinely collected by all states. Consequently questions for education policy decision-making like, “What is the total usable space devoted to K-12 education?” cannot be answered with any precision.

One potential avenue to providing this information on a national basis would be to utilize remote sensing to identify school buildings and to estimate usable space. Therefore NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of technical experts to consider how to respond to the need to estimate the physical plant for US K-12 schools, in particular the possible role for remote sensing and the resource requirements to locate or create and use these data efficiently.

The panel held several teleconferences and met in person at NCES on 19-20 February 2020.

REMOTE SENSING TO ESTIMATE U.S. K12 PHYSICAL PLANT

I. INTRODUCTION

While the National Center for Education Statistics (NCES) collects annually and maintains administrative data on all K-12 schools in the United States, there is no comparable collection of information on schools' physical plant (buildings, grounds and other infrastructure necessary for each school). Few states or smaller jurisdictions hold current comprehensive information on school facilities. Consequently there is currently no comprehensive data collection that can provide data to answer to questions such as "What is the total usable space in US K-12 school buildings?". Yet decision-makers at all levels need this kind of information as they set policy and develop facility plans.

Creating such a comprehensive data base presents problems of information sources and of scale. One avenue would rely on remote sensing imagery now that it is widely used, with street and housing images readily available on a national scale via online maps and real estate websites. How useful remote sensing imagery could be would depend on what information could be obtained and on the accuracy of source images, technical feasibility of extracting and interpreting the relevant image information, the possibility for scaling by automating the data extraction process and building the data base, and on the cost.

II. BACKGROUND

NCES had fielded an experimental effort that indicated remote sensing coupled with already available GIS school locations could be promising. Therefore NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of technical experts to advise on the possible role of remote sensing imagery in building a comprehensive facilities data base for US K-12 schools. The panel was requested to address the overall feasibility, the kinds of technical expertise required, suitable and usable imagery resources, technical problems affecting the accuracy and precision of facilities estimates, and about a structured technical approach.

For consideration of the various elements of K-12 physical plant, NCES set the following priorities (from highest to lowest):

1. Building/structure footprint
2. Square footage under roof
3. Temporary buildings – number and total space
4. Playgrounds
5. Athletic fields
6. Parking lots
7. Building condition

The objective is to create a data record for every school (possibly every campus when shared by multiple schools) in parallel with the current Common Core of Data (CCD). The CCD is a comprehensive, annual, national administrative data base of all elementary and secondary public schools and school districts. A separate comparable data base is maintained by NCES for private schools. Updating should be possible on an annual basis, but cannot depend exclusively on updated remote sensing information.

The panel's deliberations first identified possible technologies and available data resources and assessed the applicability of each to the prioritized list of elements of K-12 physical plant. Based on further discussion with the NCES Commissioner and staff about scope and resources for creating such a data base, the panel proceeded to outline a multi-stage approach.

III. METHODOLOGIES AND DATA SOURCES

To achieve the most basic goals using satellite imagery, the following data are required: identification of buildings from the images, accurate representation of the footprint of each building (in a GIS polygon) and a means of linking the building polygons, images and summary information to the CCD schools database.

It rapidly becomes clear that while remote sensing can provide the core data base, a successful plan will need to be multi-stage and also multi-modal. Validation of software for creating and curating the data base will require ground-truthing via either direct observation or existing administrative data. Resolution of inconsistent or uninterpretable image data will require "human eyes." Updating on an annual basis may also depend on other additional input, especially if imagery/image analysis is updated on a less frequent basis.

The key to this project is that the latitude and longitude of each school location is already part of current data bases such as the CCD. Consequently remote sensing can be focused in a small area if not pinpointed for the majority of K-12 schools. Exceptions will occur when some school facilities are not co-located with the school, such as sports fields either off-site or shared with other nearby schools. Also, where a school occupies only a portion of a building (e.g., a church wing, or a floor of a high-rise) other modes of data collection will be required. Especially in dense urban settings, the GIS location may not be precise enough to distinguish between adjacent buildings and cannot identify what part of a building contains the school so that additional information will be required. However, it is anticipated that the vast majority of schools will be successfully identified so that the usable space "under roof" can be estimated.

Image Requirements

The basic requirements for a usable image are clarity, current data, and an accurate reference location. Since geolocation is only an identified point, image-based interpretation depends on first identifying the relevant area surrounding that point. Errors can be catalogued as those of mistaken inclusion and those of mistaken exclusion.

The area to be searched for school buildings needs to be defined. Boundaries could be defined heuristically starting from the GIS location and using a fixed distance or shape plus any specified edges (such as a road), also possibly taking into account between-building distance. Then, computational rules would create automated boundary definitions individually for each school. Depending on the rule, problems of boundary inaccuracy would lead to mistaken exclusion when the school campus sprawls or when some school buildings lie outside the boundary (e.g., across a street). Conversely, expansive boundaries or close inter-building distances could lead to mistaken inclusion of non-school buildings.

Note that Google has invested great effort and has succeeded over time in segmenting most buildings in the US with identification of schools by location but not necessarily identification of individual buildings comprising a school.

Alternatively when other information is available, boundaries for the area can also be defined by other map data or administrative records, such as parcel boundaries. Problems with this method include that school campuses may cover multiple parcels and/or multiple schools may share a single (or multiple connected) parcel. Mixed use of buildings in an identified parcel, for example situating school district administrative offices on a shared parcel, would create an inclusion error. Distributed school facilities (across discontiguous parcels) carry the same potential for mistaken exclusion as they do for heuristic boundary definitions.

Within the bounded area, structure identification and measurement require a clear image of each building perimeter. School buildings are generally relatively easy to outline because they tend to be situated with adjacent parking area or other open space, and are usually not obscured by trees and shrubbery. Temporary classrooms can be identified by size, since they are typically the size of a double-wide trailer, about 900 square feet.

Once all relevant buildings have been identified, a set of rules or algorithms is needed to identify the space to be used to compute the school's square footage. "School building" may be defined broadly as all physical spaces within the school or narrowly as only potential teaching spaces, excluding administrative offices, theaters, gyms, cafeterias, etc. Since remote sensing gives no indication of the use of the structure beyond what is obvious from its shape, the former broad definition would be less error prone. Regardless of the definition used, it must be applied consistently across the entire project area.

Potential errors of the image analysis are inclusion of non-building space (e.g., courtyard) within the identified building perimeter and mistaking closely-placed temporary structures for a single permanent building. Filters for minimum sizes for buildings/temporary structures used as teaching space can eliminate obvious non-classroom functions (e.g., portable classroom vs. athletic shed). It is inevitable that a small percentage of school buildings in urban and non-urban settings will not be amenable to accurate automated definition of their footprints. These will require human intervention or alternative data sources.

Available Technologies

Two remote sensing technologies could be considered. Technical differences between these two mean that data are captured in different ways, thereby creating different kinds of data bases and enabling different kinds of spatial estimates. Software is not transferable; nor is a shared approach to data interpretation possible. Therefore once a choice is made, changing the technology would require starting *de novo* to develop software for data analysis and space estimation.

Remote Sensing Imagery

Satellite images, familiar from "street scene" maps, are 2-dimensional as are photographic images generally. Building measurements are effectively limited to footprints, with accuracy depending on how completely the outer walls are visible. Satellite images are subject to problems with cloud cover or weather conditions; and quality also depends on the sun's angle (time of day). The most accurate building footprints are taken from a direct overhead angle which eliminates shadows.

Object recognition software using satellite images is already available for a variety of purposes. For buildings, work is ongoing by both government and industry to upgrade software, improve accuracy and develop specialized software for specific purposes.

LIDAR (LIght Detection And Ranging)

Like RADAR (RAdio Detection And Ranging) which bounces radio waves off surfaces and records distances based on the return time of the bounced wave. Using a pulsed laser beam from aircraft, LIDAR "pings" the target surface using light waves to record distances for creating a 3-dimensional map. Thus LIDAR can produce accurate shape and volume data. Further, again like RADAR, LIDAR is unaffected by weather conditions.

The use of LIDAR data for object recognition and measurement using LIDAR has for some time been undertaken by the Department of Defense. Now it is widely under development for a variety of commercial uses as well. With respect to the specific use for building recognition and space estimation, work with LIDAR data is still in late-stage research.

The USGS (US Geologic Survey) expects 3DEP (3-D Elevation Project, a public-private partnership) to have mapped the entire US using LIDAR by the end of 2023. The primary uses of these data are for land surface information, including flood-risk management, precision agriculture, infrastructure projects, natural resources management, especially to help communities cope with natural hazards such as floods and landslides. Consequently building information will not be singled out for identification or for measurement as part of 3DEP.

Other Methodologies

Remote sensing may provide the core information, but other sources will be required as well for direct information. In addition to estimation of physical space for schools where remote sensing does not provide relevant information (e.g., partial use of a building), resolving data inconsistencies and verifying unusual boundaries will require auxiliary sources. Any of these additional data sources must include a school identifier (e.g., address, latitude/longitude, school i.d. number) that will enable automated linkage to the image and CCD data. To be most useful, these data should be collected in a consistent, accurate and timely manner across the entire U.S. (at least in areas where schools are located).

Vetting the process of interpretation of remote sensing data, and defining algorithms for automated diagnostics and/or corrections for common problems must be part of at least the initial (pilot) project and will require "ground truth" from an accurate auxiliary source.

Given a ground truth source from the initial project stage, the data analysis and modeling will require statistical and AI methodology. Aberrations in buildings identified and in square footage need to be classified by type, by need for visual review. Automated flags (e.g., too many students or too few for the footprint square footage) may be logic-based diagnostics. Other flags (included building identification or boundary mistakes) may be created using AI or statistical methods using ground truth observations to model local/regional practices in school architecture and building-boundary spatial relationships. In addition to identifying schools for review, AI and statistical methods could provide model-based corrections or imputations.

On a continuing basis, updating to improve the quality of the data base will have to rely on additional sources of information since refreshing remote sensing images is usually on a multi-year cycle while the roster of school buildings in operation changes annually.

Of particular concern is the augmentation from “footprint” to total “under roof” square footage. Essentially, the number of floors of usable space is the missing piece of information that cannot be obtained from remote sensing images but that would be provided by LIDAR. Alternative sources of each building’s number of floors would allow the conversion.

Administrative Records

No comprehensive list of schools’ “space under roof” exists on a national scale nor for most states or even school districts. However, there are a few jurisdictions for which reasonably current records do exist. Once remote sensing data is available and in usable form for estimating either footprint or total space under roof, records from these jurisdictions could be used for validating the computational algorithms. A second important use of available, accurate records is to identify different kinds of errors in estimates from remote sensing in order to create algorithms for their detection and potentially for correction. Potentially useful records are parcel data and other property records that define each legal land unit. An estimated 90% of counties have parcel records although these may not all be digitized and searchable.

A formal survey or even the addition of a question to an existing national survey would require following federal guidelines and extensive, time-consuming, and possibly unsuccessful procedures for justification and approval. Thus instituting a new survey or a new question on a mandated survey was not considered further.

A research survey to evaluate the data from remote sensing might be authorized as part of vetting the data capture/space estimation process; and additional open, public information might be gathered. Such a research objective might be achievable through a one-time addition of a question to an ongoing data collection that would allow a visual record taken onsite by survey or assessment personnel or school staff. For example, the number of floors or each identified school building might be recorded by an observer or a picture (e.g., cell phone or comparable).

Image Data Sources

Image data is available from public sources, from commercial sources such as Microsoft or Google and smaller companies that customize from primary sources and also from federal sources.

Data quality and unknown information timeliness make the public information useful for a first consideration of project feasibility but less desirable for further use. Comparably comprehensive LIDAR data bases do not as yet exist but may be on the horizon within several years.

Major mapping software companies are continually expanding and updating their products, including both image data and image interpretation software packages and outputs.

On contract from FEMA (Federal Emergency Management Agency), the Oak Ridge National Laboratories are creating a comprehensive national federal image-data base of structures with a 5-year refresh cycle. Complete data sets including images showing, coordinates for calculations of building footprints and square footage have been completed for several states. Image data sets for the remaining states are to be completed by the fall of 2021.

In 2016 FEMA contracted with the US Geologic Service for the management of the 3DEP (3D Elevation Program), as a public-private partnership program to build a national high-resolution elevation data set, essentially to create a complete 3-D map of the United States. Data acquisition is scheduled to be complete by the end of 2023.

Both commercial and federal remote sensing images are keyed to geographic locations and can integrate with other data bases (on various scales) including real estate records, US Census and other federal data. Thus, for example, real estate parcel coordinates could be a source for school property boundaries. The ease of accomplishing this depends on the extent to which parcel geolocations are in standard (preferably common) forms and are linkable to images. Data linkage is rarely simple, even with GIS locations. In this case absolute exactness may not be necessary in specifying the boundaries since buildings are relatively large and the building identification algorithms can allow a margin around the stated boundaries.

IV. MULTIPLE-MODE APPROACH AND PLAN

Overview

To be feasible, the scope of the project will have to be encompassed in several stages. In the context of such a multi-stage project, the roles that remote sensing can and cannot contribute are considered. Then a framework for the project can be structured with integration of multiple data sources. Finally, potential solutions to common solvable problems can be suggested; and truly difficult or unique problems can be set aside for individual consideration.

Five immediate assumptions:

- Feasibility, for now, requires the use of 2-dimensional images to obtain building footprints.
- Extension to 3-dimensions will be deferred to a later stage and/or to integration of data from other sources.
- Precision requirements for estimation of school structure “under roof” are only moderate and can be achieved; but measurement uncertainty and frequencies of common errors must be known.
- Using existing boundaries – property parcels – eliminates the need for elaborate algorithms to determine where to draw the line around school property.

- Updating will have to rely on other methodologies to improve accuracy as well as to account for school structures undergoing change or going into and out of use as classroom/teaching space between image revisions.

The remote image data can supply for each building the footprint dimensions and square footage, and also the location relative to parcel boundary. This makes it possible to define automatically a "campus" that includes one or more parcels to include all contiguous school property. This campus becomes the basic spatial unit for creating the inventory of school buildings. In addition, these data make it possible to filter out structures too small either to be temporary classrooms or to serve as permanent classroom buildings.

Remote image data cannot supply information about height or number of floors of usable space. Neither can it supply the primary use of a building; in particular it cannot specify the particular school for each building on a multi-school campus. Nor can it supply information about multi-use buildings, for example a high-rise building with a school on one floor or a school building that also houses school district administrative functions.

Most importantly, remote image data cannot by itself supply information about its own inaccuracies. Alternative methods must be used to identify frequencies and remedies for common (and uncommon) problems such as the inclusion of an open courtyard in a building's footprint or the aggregation of side-by-side temporary structures into an apparent single permanent building.

Thus the vetting of the estimation process for a pilot set of states is a critical part of the initial phase of the project. Comparison of the estimates from the remote sensing data with "ground truth" in the form of reliable administrative records, direct verification from school authorities, photographic/actual observation ("human eyes") will be the basis for creating diagnostics, software refinements, and an assessment of the achievable accuracy.

Following detailed analysis of the pilot states' data and software revisions including incorporation of diagnostics, this template can be applied to data from other states as FEMA completes the national (continental) inventory of structures.

Elements of School Facilities Data Record

Key elements of the public record for a school will need to allow for updating and must include at least:

- School name and address
- Date of record
    - Last confirmation of accuracy
    - Most recent update
- School information:
    - Geo-location (latitude/longitude of a central point)
    - Grades
    - Shared campus?
- Campus information:
    - Parcel number (s)
    - Image with Buildings labelled (A, B, C, . . .)
        - Date of Image

- Building A information:
    - Structure status (intact as in image, changed by addition or renovation, demolished)
        - Date of most recent update/correction/revision
    - Usage status (school, other school district building, non-school, decommissioned)
        - Date of status
    - Grades housed OR Alternative usage (e.g., district administration)
        - Source of information/Level of documentation (coded: school official, administrative record, confirmed /documented/unconfirmed volunteered information)
        - Date of information
    - Footprint (sq. ft.) from Image
        - Date of image
    - Footprint (sq. ft.) from other source
        - Source/Documentation (imputation included)
        - Date of information
    - Permanent or temporary
        - Number of buildings encompassed
            - If more than one building, then number or % space used for teaching/learning
        - Source/Documentation
    - Number of floors
        - Standard or Mixed heights (e.g., part of building is 2-story gym)
            - Net number of floors – for calculation of total square footage under roof
        - Source/Documentation (imputation included)
    - Total sq ft under roof (standard calculation or authorized information)
        - Source/Documentation
        - Date of information
    - Total sq. ft. under roof (modified calculation with all updates included)
        - Source/Documentation (imputation included)
        - Date of information
- Building B information:
    - As for Building A

Maintaining this public record relies on also maintaining a complete database of primary information such as the basic image data (e.g., vectors, coordinates in original form). In addition a database is needed with the update history of updates for all schools, active and decommissioned as buildings often come back into use, e.g., after renovation.

Project Plan in Stages

A four-stage plan for a national scale project would begin with an intensive development phase using both image and accurate administrative record data for a small number of states, ideally from different US regions. During this phase, the software would be created to define campuses and extract building (footprint) information within these boundaries, filtering out irrelevant structures. Both comprehensive visual checking and comparison to administrative records would be used to find inaccuracies and anomalies. Once identified, these could be characterized with respect to magnitude, frequency and potential for automatic detection. Diagnostics might be image-based or related to school characteristics, in either case leading to review and possibly adjustment. (A more detailed outline of the technical process appears in Appendix C.)

Initial versions of both a public data base and the complete data base of primary source information (vector data and images) with full information for reference and computations would be created. Also during the first stage, determinations of criteria would be determined for the applicability of image-based estimation of school square footage, either footprint or total under roof. Classification of buildings into permanent or temporary structures could by done by assuming that any rectangular structure that is less than 1000 square feet is a temporary (portable) classroom.

With the completion of a fully operational system for the pilot states, the second stage would comprise the extension to the remaining states. Some adaptations, including modified diagnostics and automated corrections, would be anticipated due to different architecture and construction practices in different regions and types of locales. The vetting process would be less intensive, likely to be conducted on a sample basis.

The final component to the second stage would be the full implementation of the system incorporating a non-image-based plan for updating, in particular addition of information between image-cycle revisions, and information on numbers of floors for school structures as well as structures not used for teaching/ learning. Updates would be analyzed, on occasion leading to revisions of diagnostics, automated corrections and model-based imputations/adjustments.

The third stage to scale up to 3 dimensions would be based on consideration of the efficiencies of using alternative remote sensing data (e.g., LIDAR data), combining administrative or other alternative data sources to update inaccuracies. Depending on the approach to 3-dimensional estimation, the third and fourth stages of the project overall might be reversed.

The fourth stage would address the non-building elements of schools' physical plant: playgrounds, athletic fields, parking lots, etc. Some of these tasks may be relatively easier (e.g., baseball fields, parking lots), particularly if software has already been developed, commercially or otherwise.

Anomalies, Corrections and Updating

Anomalies and Errors

With well-chosen, current data sources, errors such as inaccuracies in administrative boundary specifications are likely to be without effect because they are too small to impede detection of buildings within the bounded space. Gross data extraction problems such as failure to recognize a building within the image or failure to detect the proximity of a building to a parcel boundary (necessitating consideration of including the adjoining parcel in the campus) can be assumed to be rare. In any case, analysis of the likelihood of this kind of extraction problem, its detection and correction would normally be a part of the process of building the data base.

Other problems at this stage would include incorrectly drawn footprints or mis-measurements. Small deviations, for example inclusion of a small porch or portico, are likely to fall within allowable measurement variation for the intended and attainable precision of total square footage estimate. However, the inclusion of a large open courtyard within a building footprint or the circumscribing of a perimeter of grouped permanent buildings is not ignorable. Failure to detect side-by-side temporary buildings as distinct might affect square footage minimally but would impact estimates of permanent space and numbers of temporary structures.

Interpretation errors could include misidentification of a building (belonging/not belonging to the school or permanent/temporary or whether teaching/learning space or not). This teaching/learning space (measured as “area under roof”) should include all classrooms and hallways - whether temporary or permanent - other facilities used by students (e.g., gyms, auditoriums, practice rooms, bathrooms, etc.), plus school administrative offices, etc. Square footage should exclude maintenance and storage facilities, janitorial closets and mechanicals rooms, and also exclude school district offices and administrative space. Identification of multi-story buildings is the most problematic aspect when footprint does not correspond to square footage under roof. Volume is not a simple and complete answer as tall buildings can house athletic facilities (e.g., gymnasiums, swimming pools) or theaters but be primarily single floor buildings while tall classroom buildings have multiple floors. However volume is a useful additional measure for considering building upkeep and maintenance costs.

Identification and Correction

Identification, classification and evaluation of footprint data errors comes from comparison of the set of footprints for each campus with a “footprint standard.” That standard is created during Stage 1 by the data base of “human eyes” structure outlines created for the schools in these pilot states. From these data, the kinds of errors can be inventoried and measurement uncertainties quantified: inclusions/exclusions (of buildings and of spaces), mis-measurements and/or miscalculations of usable square footage. Decisions based on frequency and impact of these errors can be taken for corrections, whether to use an algorithmic correction, to impute values, to refer for corroboration from either administrative data or “human eyes” review or to ignore. (For later stages, the exhaustive manual examination would likely be reduced to sampling with the goal of expanded definitions of diagnostics to suit local architecture and siting practices.)

Detection of most interpretation errors would rely on comparison with footprint data or summary calculations from footprint data with alternative data sources. CCD records provide information on size of student population (by grade level), per pupil expenditure and other school characteristics. Relevant information might also include state or local regulations for space/student, age of building, median income for census tract and other information in a federal or public data base. Architectural norms for a region can be analyzed from the administrative records for Stage 1 and from a sample of schools for Stage 2. Total square footage under roof could be modeled based on all useful, relevant attributes to create a range of most likely values to use in defining a diagnostic and/or in imputation. Decision options are once again: to use an algorithmic correction, to impute values, to refer for corroboration from either administrative data or "human eyes" review or to ignore.

Updating

Since schools regularly expand, contract, reassign, open and close facilities, an annual updating system is needed outside the (5-year) revision cycle for image data. Non-image data or imputation will be required to account for all footprint changes. Changes in facilities for an existing school will go undetected unless a change in student population or other key parameters is diagnostic. Absent a requirement for submitting this facilities information or willing cooperation of a school official, imputation may give the best estimate for a new school on a site not previously occupied by another school.

Cooperation in validating and in improving the information in the public data base can be sought from school officials, from teachers/students, from onsite observers. Administrative confirmation can be sought for changes, but need not be required provided it is clear how authoritative the data are known to be.

Annually an invitation can be issued to each principal, school district superintendent and state department of education to review the current data and provide updates. Submission of changes might be allowed at any time. (Preferably this updating process should be electronic with an opportunity for official authorization.)

Visitors onsite could have the opportunity to do something as simple as visually assess the number of floors of each building and enter that information (perhaps with photo documentation) into the data base update system.

"Citizen Geography" at the level of interested citizens or by way of geography modules for students could provide either simple (number of floors per building) or more comprehensive (estimated square footage with allowances for building of mixed height floors) information.

Resources Required

The bulk of the investment would be in creating the databases, related software and the system for updating, followed by the initial national-scale implementation. Depending on the availability and accuracy of parcel data and the efficiency of imputation/adjustment for multi-story buildings, the task may need to be expanded, for example to use AI to develop software to define boundaries, detect anomalies and define adjustments.

Once created, the rolling revisions to the image data on a 5-year cycle would necessitate only a rerunning of the extraction software, the computational software and a review of anomalies. Between image revisions done annually on a rotating set of states, other methodologies would provide the updating. Thus unless/until a different modality is later adopted (e.g., LIDAR) expenditures would be concentrated at the outset, particularly during the first stage.

Given availability of remote sensing images for several states with accurate administrative data, stage one could be accomplished expeditiously, likely requiring approximately 1.0 - 1.5 man-years of highly skilled technical work. This would logically come from a team since the technical requirements for different parts of the project are quite specific and definitely diverse, emphatically so when modeling to create diagnostics and/or adjustments is considered. Even computer languages differ among these tasks. The tasks of data extraction from vector data files, data integration (file linkage), computation and analysis to detect inaccuracies and to define diagnostics also model-based corrections and imputation algorithms, public and primary reference file construction and file management including updating. Skills required include computer science (for image extraction, analysis and verification including human validation),statistics, data science (i.e., AI, machine learning, and high-dimensional data analysis), computer science (for data integration and file structures).

V. FINDINGS AND RECOMMENDATIONS

The charge to the panel was to determine the feasibility of estimation of school physical plant, including both structures and dedicated exterior spaces, and to evaluate the potential for using remote sensing data to produce these estimates. The panel found that remote sensing data could be used for a reduced set of objectives and provided an outline for one way in which this task could be accomplished.

Principal Findings

Remote sensing imagery can be used to create the foundation and primary information data base for a national inventory of K-12 school space for teaching/learning.

With current remote sensing technology, estimation of the square footage for school structure *footprints* is possible on a national scale. At present, estimation of total square footage "under roof" requires incorporation of information from additional sources.

Classification of buildings as permanent or temporary structures may be achieved by categorization based on square footage. Estimation of visually distinctive athletic fields and of parking lots may be possible as a separate task using software already developed or in development either from commercial (athletic fields) or from federal sources (parking lots). Playgrounds and open athletic fields may not be distinguishable from other open space. Remote sensing technology is not suited to providing information on either building condition or internal utilization of building space.

Special cases requiring additional or alternative approaches include residential campuses, partial building occupancy (for example schools situated in urban high-rises) and subterranean structures.

A system for updating school facilities information is essential as schools' physical plants are continually undergoing change. Annual opportunity for updating is needed but must rely on alternative methodology as remote sensing information is only revised periodically (e.g., a 5-year cycle of re-imaging).

Recommendations

1. Estimation of US K-12 physical plant to be undertaken in stages.

Stage 1 to be limited to comprehensive data for a pilot set of states with both administrative information on buildings for each school (footprint, square footage under roof) and parcel data in a usable form to be integrated with remote sensing imagery. Stage 1 to focus on definition of a data base (and associate primary data base) for two-dimensional information (footprints), and to include validation, development of diagnostics and adjustments, with a view to detection of multi-story buildings, verification and consequent adjustments.

Stage 2 to scale up to national level with incorporation of updating processes.

Following completion of Stages 1 and 2, Stages 3 and 4 to be reconsidered with respect to priorities and with respect to scaling up from 2-dimensions to 3-dimensions for primary data and the efficient estimation for multi-story buildings and with respect to available options for detection and inventory of non-structure facilities including athletic fields, playgrounds and parking lots.

2. Use of existing data bases and existing software to be maximized. In particular, the federal data base (USA Structures – being developed at the request of FEMA), and data bases with parcel boundaries geo-inscribed can avoid redevelopment of software and instability of data sources over time.

3. A formal system for updating on an annual basis to be built, independent of remote sensing, to reflect changes in school facilities between revisions to remote sensing imagery data.

APPENDICES

Appendix A: Agenda

IES Institute of Education Sciences

Expert Panel on Using Remote Sensing to Estimate U.S. K12 Physical Plant Meeting

February 19–20, 2020 | Washington, DC

Agenda
Room 7080

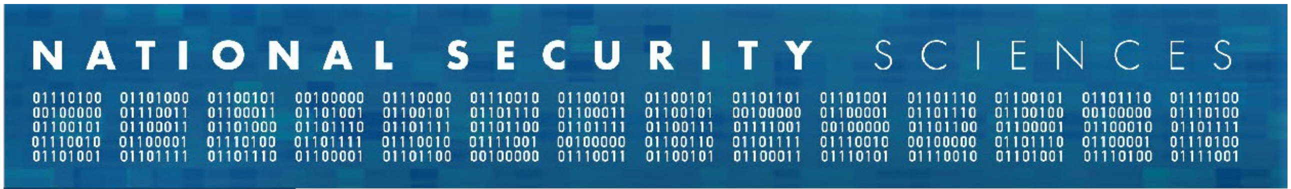
Wednesday, February 19, 2020

8:30 a.m.	Arrival and Building Security
9:00 a.m. – 9:15 a.m.	Welcome and Introductions
9:15 a.m. – 10:45 a.m.	Review of charge and Presentation of work at NCES
10:45 a.m. – 12:15 p.m.	Questions from the Panel and Discussion
12:15 p.m. – 1:15 p.m.	Lunch (on you own)
1:15 p.m. – 4:30 p.m.	Panel Executive Working Session
4:30 p.m. – 5:30p.m.	Panel Clarification Requests for NCES (as needed)
5:50 p.m.	Adjourn

Thursday February 20, 2020

8:30 a.m.	Arrival and Building Security
9:00 a.m. - 10:30 a.m.	Panel Executive Session
10:30 a.m. – 11:45 a.m.	NCES Responses to Panel Requests (if useful)
11:45 a.m. – 12:45 p.m.	Panel Executive Working Session
12:45 p.m. – 2:00 p.m.	Move to Room 7085 Before Lunch (on your own) and return to 7085 any time before 2:00 p.m.
2:00 p.m. – 3:45 p.m.	Panel Executive Working Session
3:45 p.m. – 5:00 p.m.	Summary Session with NCES
5:00 p.m.	Adjourn

Appendix B: Data Source – USA Structures Project



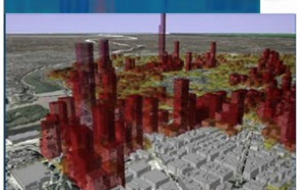
Contacts

Melanie Laverdiere
Research Scientist
Remote Sensing

865.574.7239
laverdieml@ornl.gov

Mark Tuttle
Project Manager
Human Dynamics

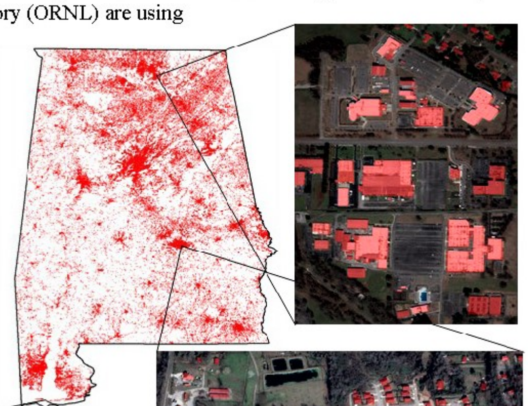
865.241.4150
tuttlema@ornl.gov



USA Structures

Using AI and HPC to Build a National Structures Inventory

Each year, floods kill more people and cause more economic damage than any other natural disaster. In 2016, the U.S. experienced 32 major disasters and six emergency declarations involving floods. The Department of Homeland Security (DHS) Science and Technology Directorate (S&T), Federal Emergency Management Agency (FEMA), and the National Geospatial Intelligence Agency (NGA) are partnering to build an accurate, up-to-date, and complete US structures inventory with detailed attribution to better understand areas at risk and the potential flood impacts within these areas. In support of this effort, researchers at Oak Ridge National Laboratory (ORNL) are using satellite imagery, coupled with high performance computing (HPC) and artificial intelligence (AI), to automate a comprehensive inventory of at-risk buildings and infrastructure. This new data will allow companies to properly insure private and commercial property, and flood planners to design more effective flood-protection strategies.



Solution

ORNL is developing algorithms based on advanced deep learning techniques that can reliably and consistently detect buildings of varying land uses such as residential, commercial and industrial from aerial and satellite imagery. The key innovation is to apply advanced computer vision technology to detect physical structures in the built environment.

Because the process uses AI and super-computing techniques with algorithms that actually improve in accuracy the more they are used, inventories can be kept up-to-date much faster and at a far lower cost than before.

This data set, coupled with detailed risk information about local flood hazards, will help first responders predict where damages are most likely to occur, how extensive damages are likely to be and where agencies should deploy resources for greater impact. Because the data will be linked to information on property ownership, tax assessments and current insurance, the inventory will also help FEMA and local planners target flood mitigation investments more accurately and improve FEMA's related public assistance programs. Other beneficiaries will be the Army Corps of Engineers' capital investments in waterways and the National Weather Service's flood forecast modeling.

The project has already developed and tuned the initial algorithms needed to detect all buildings and their features in a pilot area. The current phase will scale the process to cover the entire country including all territories, while continuing to refine the algorithms. The goal is to complete the first inventory of the entire U.S. by the fall of 2021.

Date: February 5, 2020

OAK RIDGE
National Laboratory

Appendix C: Multiple Stage Approach and Plan

Goals: *Initially: Estimate square footage under roof of teaching/learning space for K-12 US schools*
Ultimately: Complete estimation of all physical spaces used in operating K-12 US schools

Remote Sensing Limitations

- Space utilization (Usage of space: administrative, classrooms, storage, empty etc.)
- Space allocation for co-located schools (sharing)
- Building Conditions
- School space within other buildings

"Ground Truth" Resources

- Accurate administrative (state) record
- "Human eyes" assessment of images
- Onsite observation

STAGE 1 – Develop, Implement and Evaluate Pilot Project to Estimate Footprint Square Footage for Each Structure and Each School

- Temporary and permanent buildings with usable classroom space
- 4-6 Pilot states with complete remote sensing data and definitive state data
    - MA, AR, with VA, OR, and Southwest state if state data are available

Step 1 – Data File Definition: Assemble Data Sources

- Parcel data to define campus boundaries
    - Vector data
- Common Core Data (CCD)
    - All schools in pilot states
    - School geo-location data
    - School descriptive information
        - Grades
        - Student population
        - Other relevant attributes (useful in relationship to space required)
- Oak Ridge National Laboratories (ORNL) - USA structures program for Federal Emergency Management Agency (FEMA)
    - Building/structure files for pilot states
    - Vector data
- State Information
    - School campus - Parcel(s)
    - Building/structure information (individual building)
        - School within campus – Grades
        - Permanent or Temporary
        - Footprint
        - Number of Floors
        - Total square footage under roof
        - Current usage

- Statutes and practical space requirements (by grade)
- Tabular data

- Imagery Collection (human validation; chip ID)
  - Independent assessment for all schools in each pilot state
  - Digital globe imagery

Step 2 – Data File Definition: Logic Checks and Diagnostics (By State)

- Analyze CCD information to derive a set of logic checks (examples)
  - Size range (total sq. ft. under roof)
    - minimum size: based on student population for grade levels
    - maximum size: based on %-ile (90? 95?) of state record data for grade levels
  - Predicted student population for footprint – comparison to state record
    - Single story
    - two (also for more) stories
  - Predicted number of stories by school attribute (e.g., location, grade level, structure type)
- Implement filters to flag building diagnostics then validate

Step 3 – Data File Definition: Create Filtered List of School Structures (By State)

- Filter out buildings based on parcel data
  - Define campuses surrounding school geo-location (point)
    - If a vector for a building intersects (or lies close to) a vector of a parcel
      - Adjoin adjacent parcel to create campus
      - Check for other schools on adjacent parcels - determine whether these are separate or decide to create a multi-school campus
  - Address all flags and all failure of logic checks
    - Manually adjust campus boundary
    - Manually designate structure grade level, usage or other attribute
- Categorize buildings based on image
  - Threshold size minimum of approximately 700 sq. ft. (900 sq. ft. is common size for temporary structures)
  - Threshold size minimum of approximately 1200 sq. ft. for permanent structure
    - Evaluate for possible mistaken combining of several temporary structure into one
  - Categorize structure by location – e.g., structure adjacent to track/stadium

Step 4 – Data File Definition: Create data file from remote sensing images for pilot states

- Build computational software
- Output
  - Store vector information by building
  - Total square footage per building
  - Total permanent square footage by school or campus
  - Total square footage of temporary buildings
  - Number of temporary buildings
  - Number of buildings per campus parcel also by individual parcel

Step 5 – Validation: Comparison to Existing Definitive Data

- State Record Information
    - Total square footage per building
    - Total permanent square footage by school or campus
    - Total square footage of temporary buildings
    - Number of temporary buildings
    - Number of buildings per campus parcel also by individual parcel
- Visual review using GIS framework with overlays (randomly chosen subset)
    - Overlay imagery and vector building data
    - Tools that allow manual input
        - Manually change the AI Segmentation

Step 6 – Validation: Analysis of Errors and Uncertainties

- Manual input ("human eyes") plus definitive records as "ground truth"
- Choose the right metrics to compute the error
- Inventory errors
    - Type
    - Magnitude
    - Frequency
- Validation process
    - Model error probability
        - Based on CCD information
        - Based on image information
        - Based on diagnostics (type, value if numeric),
    - Determine when automatic correction algorithms are efficient (for which errors)
- Technical method uncertainties from ORNL data

STAGE 2 – Extend Implementation to All States

Step 1 – Assemble Data

- Load CCD Information including geo-location
    - Exclude schools within other buildings/structures (as for pilot)
- Load parcel data
    - Suitable Parcel database available for all states
    - Alternative strategy defined
    - Flag parcel if point is not inside of parcel
- Exclude schools within other buildings/structures (as for pilot)
- Load ORNL – FEMA USA structures data base with updates

Step 2 – Create Data Base

- Follow workflow established for Stage 1
- Modify or create alternate diagnostics to adapt to regional/state differences
- Abbreviate comprehensive validation by sampling

Step 3 – Develop System for Updating, Corrections and Sustainability

- Human validation
    - Metadata analyses with uncertainty analyses - annually
        - ORNL data
        - Observer updates
        - Administrative corrections/changes
    - Managed data base
- Alternative data sources to remote sensing data
    - Administrative additions (e.g., new schools/ discontinued schools)
    - Volunteered information – confirmed/unconfirmed
    - Survey - official or informal
    - Web-based data input system
- Rolling 5 year updated with image data sources as available

STAGE 3 - Move to 3-D Technology for 3-D Computation (Height, Volume)

- Development, implementation and validation process to follow Stage 1
    - Pilot work on 4-6 states
    - Scale up to all states
- Technology
    - LIDAR
    - Google Street view
    - Satellite imagery - new or existing algorithms / USGS 3DEP expected in 2023
- Complications
    - Previous technical work on 2-D images not transferable
    - New model development for addressing 3D data representation
- Alternative or supplemental solutions
    - New data sources
        - Surveys
        - Observational data
        - Crowd-sourcing

STAGE 4 – Scaling Up: Additional Facilities

- Priorities and work scope to be determined
    - Athletic fields
    - Parking lots
    - Playgrounds
    - Additional resources
        - Data bases such as National Storm Shelter System
        - Extraction software for football stadia, baseball fields, tennis courts

Appendix D: Expert Panel Biosketches

Peter Bajcsy, PhD

Title: Project Lead, Software and Systems Division, Information Technology Laboratory, National Institute of Standards and Technologies

Dr. Peter Bajcsy received his Ph.D. in Electrical and Computer Engineering in 1997 from the University of Illinois at Urbana-Champaign (UIUC) and a M.S. in Electrical and Computer Engineering in 1994 from the University of Pennsylvania (UPENN). He worked for machine vision, government contracting, and research and educational institutions before joining the National Institute of Standards and Technology (NIST) in 2011. At NIST, he has been leading a project focusing on the application of computational science in metrology at very large scales. Peter's area of research is large-scale image-based analyses and syntheses using mathematical, statistical and computational models while leveraging computer science foundations for image processing, machine learning, computer vision, and pattern recognition. He has co-authored more than 40 papers in peer reviewed journals, about 100 conference papers, and 11 books or book chapters.

Brady Cline, BS

Title: Business Development, GeoAI, ESRI

Mr. Cline has over 20 years of applying software and data to address new and existing challenges in the US DoD, federal, and international ministries/governments for intelligence, security, and disaster response. The past ten years have been focused on ensuring spatial data is enabled for rapidly building actionable intelligence and course of action analysis. Mr. Cline is currently supporting Esri's Business Development for GeoAI, as a cross-sector resource. Prior to joining Esri, He was on the management team at SpaceKnow, leading global sales, business development, and marketing. In that role his focus was broadened to include commercial and financial industries, and enabling them to utilize best in breed Artificial Intelligence tools for spatial data extraction, that previously seemed only available within National Intelligence.

Joe Chalfoun, PhD

Title: Project Lead, Software and Systems Division, Information Technology Laboratory, National Institute of Standards and Technologies

Dr. Joe Chalfoun received his Doctoral degree in mechanical engineering and bio-robotics from the University of Versailles, France, in 2005. After working on the nuclear robotic manipulator at the Atomic Energy Commission (CEA) in France, he joined the National Institute of Standard and Technology (NIST) as a research scientist since 2007. Dr. Chalfoun's research interest is in medical robotics field, mainly in cell biology applied to large dataset experiments. He likes to develop cutting edge technologies in computer algorithms to help expand biologists' capabilities to conduct experiments and perform measurement sciences otherwise not reachable by existing methods. Dr. Chalfoun's focused research areas are in Image analysis, artificial intelligence, modeling, data mining and pattern recognition, microscopy control and benchmarking, Robotics, and automation.

Shaowen Wang, PhD

Title: Professor & Department Head, Richard and Margaret Romano Professorial Scholar, Department of Geography and Geographic Information Science; Affiliate Professor, Computer Science, Information Sciences, Urban & Regional Planning; and Founding Director, CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign

Shaowen Wang is a Professor and Head of the Department of Geography and Geographic Information Science; Richard and Margaret Romano Professorial Scholar; and an Affiliate Professor of the Department of Computer Science, Department of Urban and Regional Planning, and School of Information Sciences at the University of Illinois at Urbana-Champaign (UIUC). He has served as Founding Director of the CyberGIS Center for Advanced Digital and Spatial Studies at UIUC since 2013. His research interests include geographic information science and systems (GIS), advanced cyberinfrastructure and cyberGIS, complex environmental and geospatial problems, computational and data sciences, and spatial analysis and modeling. His research has been actively supported by a number of U.S. government agencies (e.g., CDC, DOE, EPA, NASA, NIH, NSF, USDA, and USGS) and industry. He has published 100+ peer-reviewed papers including articles in 30+ journals. He has served as a member of the Board on Earth Sciences and Resources of the National Academies of Sciences, Engineering, and Medicine since 2015.

Melanie Laverdiere, PhD

Title: Research Scientist, Remote Sensing Group, Oak Ridge National Laboratory

Melanie is a Research Scientist in the Remote Sensing Group at Oak Ridge National Laboratory (ORNL). She received her bachelor's and master's degrees in Earth System Science with a focus in Remote Sensing from the University of Alabama in Huntsville in 2011 and 2013, respectively. She has been at ORNL since September 2013, first as a Post-Masters Research Associate and currently as a Research Scientist since January 2017. Her research experience has spanned across multiple areas including population distribution and dynamics, large scale feature extraction via high resolution airborne imagery, satellite imagery pre-processing, and accuracy assessment techniques for remote sensing applications.

Mark A. Tuttle, MA

Title: Project Manager, National Securities Emerging Technologies, Human Dynamics Group, Oak Ridge National Laboratory

Mark Tuttle is a Project Manager for Human Dynamics Group at Oak Ridge National Laboratory (ORNL). His focus is applying lessons learned from his diverse geospatial technology and multiyear, multi-participant project management background to multiple efforts currently underway within the Geographic Information Science and Technology (GIST) Group. Specifically, he serves as the Project Manager for a 10-year effort to develop a variety of critical infrastructure spatial datasets and a multi-year effort to develop the first high resolution spatial building inventory for the US. Both efforts are targeted to the national, state and local emergency preparation and response community. Cumulatively, Mr. Tuttle has been member of the GIST Group for 18 years and been an integral part of a variety of efforts ranging from coordination of the GIST group responsibilities for DOE emergency preparation and response for the Visualization and Modeling Workgroup, micro-scale modeling and simulation, transportation impacts and analysis, geospatial knowledge discovery. Mr. Tuttle previously served the State of Tennessee as the Director of GIS Services in the Office for Information Resources, Department of Finance and Administration for seven years.

Linda Williams Pickle, PhD

Title: Founder & Chief Statistician, StatNet Consulting, LLC & Adjunct Professor, Department of Geography, Penn State University

Linda has a PhD in Biostatistics from Johns Hopkins University (1977) and over 40 years' experience in the areas of disease rate mapping, geovisualization and the application of statistical models to health outcomes. Her work experience was with the National Cancer Institute, the National Center for Health Statistics, Georgetown University and as an independent consultant. She is an Elected Fellow of the American Statistical Association . She has published 3 disease rate atlases and a book titled Visualizing Data Patterns with Micromaps (Carr and Pickle, 2010) and has developed multi-level spatial statistical models, including one used to predict the annual number of new cancer cases by the American Cancer Society (Pickle, Hao et al., 2007). Geovisualization research included a series of cognitive experiments on the elements of map design. All of her research required locating or building large spatially-located databases and analyzing them in a manner that is statistically correct and that allows for rapid processing to produce results in a timely manner.

Panel Convened by the National Institute of Statistical Sciences

Nell Sedransk, PhD

Title: Director, National Institute of Statistical Sciences-DC

Dr. Nell Sedransk is the Director of the National Institute of Statistical Sciences. She is an Elected Member of the International Statistical Institute, also Elected Fellow of the American Statistical Association. She is coauthor of three technical books; and her research in both statistical theory and application appears in more than 60 scientific papers in refereed journals. The areas of her technical expertise include: design of complex experiments, Bayesian inference, spatial statistics and topological foundations for statistical theory. She has applied her expertise in statistical design and analysis of complex experiments and observational studies to a wide range of applications from physiology and medicine to engineering and sensors to social science applications in multi-observer scoring to ethical designs for clinical trials.

Ya Mo, PhD

Title: Research Fellow, National Institute of Statistical Sciences; Assistant Professor, Boise State University

Dr. Ya Mo is a research fellow at the National Institute of Statistical Sciences and an assistant professor of Curriculum, Instruction, and Foundational Studies at Boise State University. She received a dual major Ph.D. in Measurement and Quantitative Methods and Curriculum, Instruction, and Educational Policy Programs, and an M.S. in Statistics from Michigan State University, as well as an Ed.M. in TESOL from Boston University. She researches quantitative methods, psychometric measures, and survey statistics; she also applies quantitative research methods to study substantive topics in education, especially large-scale assessments.

Megan Glenn, BS

Title: Assistant, National Institute of Statistical Sciences

Megan Glenn is Assistant at the National Institute of Statistical Sciences working under the direction of Dr. Nell Sedransk on technical panels in education research, in particular conducting research and compiling data on scientific background on focus topics in education. She received her BS degree from Keene State College in New Hampshire.