Privacy-Protecting Commute Shed Study

Resubmission Date: November 15, 2002

Submitted to TRB 2003, Committee on Urban Data and Information Systems, A1D08

Word Count: 6,200

Steve Raney

Cities21.org
1487 Pitman Ave., Palo Alto, CA 94301, (650) 329-9200. steve_raney@cities21.org.

ABSTRACT

A new methodology has been developed for collecting fine-grained employee commute origination data from employers in major employment centers. Many U.S. multinational firms have adopted the European Parliament Privacy Directive 95/46/EC, the strictest privacy law to date. The methodology discussed complies with this Directive. "Fine-grained" is defined as providing sufficient resolution (approximately 1,000 feet) to assist transportation planning of individual bus stops. Directive 95/46/EC provides for "anonymization" of data to where the data subject is no longer identifiable. Individual address data is aggregated to a 1/5 by 1/5 mile grid at the employer site using commercial geographic information systems software. Once the grid data is taken from employer premises, data is combined with those of other employers, providing further anonymization.

8,200 out of approximately 20,000 worker addresses have been collected from the Stanford Research Park employment center in Palo Alto, California using this methodology. While the sprawling spatial distribution of these addresses challenges many Transportation Demand Reduction strategies, planned transit system improvements should result in a significant patronage increase.

This methodology could be automated and applied nationally by the U.S. Census Bureau as part of their Transportation Planning Package, creating higher quality transportation data for improved investment decision making, ridesharing, and transit routing.

THE SETTING

Stanford Research Park (SRP), located in Palo Alto, California, is counted among the nation and world's leading scientific research employment centers. SRP, also known as the "parent of Silicon Valley," has served as a model for numerous succeeding office parks. The 1,000-acre park boasts 161 buildings serving 160 companies, the twenty largest of which account for more than 90 percent of the park's 20,000 employees. Employers cover a broad spectrum of activities: personal computer manufacturing, medical systems, defense, bioscience, technology law, management consulting, information systems, electronic commerce, veterans' hospital, and research laboratories.

The study coincided with the Internet Bubble's collapse. Indeed, more than half of large employers initiated substantial workforce reductions during the study period.

SRP is owned by Stanford University. Office lease profits from the park subsidize university housing and academic programs. Stanford and Santa Clara County recently signed a General Use Permit enabling a five million square foot expansion of Stanford's holdings, with zero permitted "net new auto trips" resulting from this construction.

Palo Alto is a San Francisco Bay Area suburb of 60,000 residents, located between San Francisco and San Jose, at the northernmost tip of Santa Clara County. At the height of the Internet Economy, peak commute hour congestion reached record highs on the highways leading to Palo Alto. The Association of Bay Area Governments has projected 240 percent increased peak hour congestion for the entire Bay Area from 1990 to 2020. (1)

STUDY RESULTS

The originations of the 8,200 Stanford Research Park (SRP) workers reveal a sprawling commuting pattern (see Figure 1) – with some commuters suffering through 75 mile, two-hour commutes from Manteca and Tracy in California's Central Valley. Within the sprawling data, patterns emerge indicating a strong likelihood for increased transit use and carpooling. 47 percent of SRP employees live within a two-mile, "bikeable" radius of a Caltrain commuter rail station. Significant service improvements, known as "Baby Bullet," are planned in the coming years. San Francisco to Palo Alto commute time is expected to decrease from 65 to 45 (or fewer) minutes, making rail faster than single-occupancy peak-hour vehicle driving via congested Highway 101. Once at the Caltrain station adjacent to SRP, improved bus shuttle service, car sharing, and bike lockers ease the connecting "last mile" trip to the office.

In addition, 28 percent of these workers live within a five-mile radius and 49 percent live within a ten-mile radius of SRP. New "bus preference" service where buses control signals and "queue jump" using the right hand turn lane trims the bus/auto speed gap along the main arterial.

The client, City of Palo Alto Chief Transportation Officer Joe Kott, was satisfied that the study cost-effectively assisted better transportation planning: "Because of the upcoming Highway 101 Corridor Study, planned Caltrain Baby Bullet service, Stanford General Use Permit trip reductions, and ongoing VTA/SamTrans bus route planning, Palo Alto needed more accurate and more current commute data than other cities and transit agencies. We've accomplished this for a fraction of the cost it would have taken using a traditional transportation consulting

firm. Compared to recent regional transportation studies our data is twenty times more precise and our participation rates are much higher."

THE NEED FOR BETTER COMMUTE DATA

In the private sector, “know your customer” is an oft-repeated dictum. In the more budget-constrained transportation world, customer knowledge has been harder to achieve. Public sector transportation decisions are often made without sufficient information. The cost and effort to collect this data is often prohibitive. By providing more detailed and more current information, this commute shed methodology can provide improvements in national data, transportation investments, ridesharing, and transit routing.

Previous Origination-Destination studies have been conducted, but on a more coarse-grained level. The U.S. Census Bureau collects Traffic Analysis Zone (TAZ) data showing commute patterns in their Census Transportation Planning Package Journey to Work analysis. Each irregularly shaped TAZ consists of about 10,000 people. The Census Bureau surveys about ten percent of the population to collect TAZ data. Journey to Work data is expensive to collect, time consuming to process, and is only collected once every ten years. The Bay Area’s lead transit entity, the Metropolitan Transportation Commission (MTC), creates custom data packages from Journey to Work data. From the 1990 Census, MTC produced more than 25 downloadable data packages between April 1992 and January 1996. Given the ten-year lag between the Census and required time to create data packages, the results often do not reflect the current status.

In 1999, Santa Clara County’s transit authority, the Valley Transportation Authority (VTA), conducted a “Commute Service Study” collecting employee origination zip codes from major employers. Each Santa Clara County zip code represents about 10,000 people. Participation rates were about twenty percent for this study, which examined twelve major employment centers ranging in size from 17,000 to 67,000 employees. (2)

Improved National Data

As a supplement to their Journey To Work data, the U.S. Census Bureau could adopt this methodology, automate data collection, and implement it for all major U.S. employment centers. As a result, more accurate national data could be rapidly, inexpensively, and frequently collected and disseminated. Operating a national commute sheds survey every two years would alleviate problems with outdated data. Proposed implementation tactics encompass four areas: funding, employer processing burden, software/systems concerns, and national process.

Should the demand for this data be sufficient, data management could be funded via user fees rather than by the general taxes. At the local level, the cost of undertaking automation is prohibitive. On the other hand, a national program would provide the necessary scale to fund development of a system to automate data collection, dramatically reducing the cost of individual commute shed studies.

Preparing data for such a survey would be easy for employers with readily perceived benefits, ensuring high participation rates. A single Human Resources employee should be able to collect and partially aggregate (via the “last two digit zeroing” technique described later) employee addresses, and securely transmit the results to the Census Bureau system in less than 30 minutes.

Regional transportation agencies would define the major employment center geographic boundaries. Employers with more than a set threshold of employees nationally would supply a spreadsheet with employment address and aggregated home address for each employee. Data collection could be well-orchestrated nationally with sufficient publicity to complete collection in a single week. The Census Bureau could determine the mid-size to large companies to target, make contact, and track participation.

Better Transportation Investment Decision Making

Multimillion dollar transportation investment decisions are regularly made with incomplete and outdated data. Part of the City of Palo Alto’s satisfaction with the study was based on their newly obtained ability to more effectively influence regional transportation spending priorities.

A second Bay Area city, South San Francisco, is undertaking a commute shed study, creating a more current and accurate snapshot of transit patterns to argue for advantageous transit station placement as part of a new regional transportation investment. The City Manager’s Office disputes the regional project analysis because it is not perceived to reflect the city’s recent employment growth. From the city’s standpoint, a small investment in a commute shed study pales in comparison to the inefficiency that would be caused by locating the transit station inappropriately, far away from workers.

More Effective Ridesharing

The central Bay Area ridesharing organization, RIDES, works in a reactive mode, attempting to form a car or van pool when a person voluntarily expresses interest in an alternative commute. The commute shed allows proactive ridesharing marketing via bi-directional communication between the data collection entity and employers/employees. When the commute shed data set is created, origination data flows from employer to collection entity. Once the data is plotted on a map, promising neighborhoods with many commutes to the employment center can be identified. The collection entity may pass information about the promising neighborhoods to a ridesharing organization that may then communicate target zip (or zip+4) codes to employers. The employers will then look up employees in those zip codes and communicate a specific request to consider joining a carpool.

Because of sprawling suburban residential patterns, carpool formation does not typically have sufficient scale on an individual company basis. Aggregation to a 20,000-person employment center improves efficiency. Individual employers lack sufficient scale but can contact their employees. Collection entities can create sufficient scale, but cannot directly contact employees without running afoul of privacy rules. Bi-directional communication solves the scale problem for employers and maintains a layer of privacy separation between employees and collection entities.

Improved Transit Route Planning

Commute shed studies can also serve to improve bus and rail route planning. The Bay Area Transportation and Land Use Coalition represents more than 90 nonprofit and advocacy organizations in the Bay Area. In their Year 2000 "World Class Transit Report," they identify a pressing need to more accurately match transit service with commuter travel patterns: "There are 3.4 million workers that travel to job sites throughout the Bay Area, and yet there is no existing means to effectively match new bus service with workers' travel patterns. Transit agencies are forced to predict where commuters need to go, instead of being able to use actual data. Data could be collected … and be made available to transit agencies. Transit agencies in turn could glean valuable insights from a centralized database showing employee origins and destinations. Data collection would not impose an onerous burden on companies. Companies would stand to benefit as well, as … bus service well-matched to commuters' needs would increase transit ridership". (3) Fine-grained data allows transit planners to run GIS queries to count people within X feet of a stop / station, queries that were previously impossible.

Santa Clara County VTA implemented significant service modifications based on their coarse-grained 1999 commute shed study. Their analysis led to the following proposals: a) five new express bus routes to serve 1,000 new daily riders, b) adding/modifying bus stops on more than 50 percent of commute bus routes, c) new park and ride facilities to increase patronage on existing routes, and d) marketing budget modifications to emphasize specific routes with higher potential patronage. VTA's Principal Transportation Planner Chris Augenstein assisted in conceptualization of this finer-grained methodology, and, upon seeing the study results, indicated interest in studies for additional employment centers and interest in regular data updates for each employment center.

For new transit service, not only will fine-grained data be useful, but the bi-directional communication technique discussed above can also be used to obtain a "second opinion" about the effectiveness of proposed service. Data collection entities can work with employers to survey potential patrons based on their zip code.

For forecasting patronage of office park shuttle service, comprising the last leg of a bi-modal commute, fine-grained data allows grouping of potential riders in the larger commute shed based on distance to high quality commute alternatives such as carpool HOV lanes or express commuter rail. Fine-grained data also improve marketing budget decision making for these same bi-modal commutes by revealing promising neighborhoods to sell into. The forecasting technique can be extended to new shuttle technologies, such as personal rapid transit.

LIMITATIONS

This methodology is primarily useful for major employment centers of 10,000 or more employees and for "journey to school" studies for large universities. It does not aide in the understanding of non-work trips. For work trip planning questions where more detailed information about commuter characteristics (such as working hours) are required, individual surveys are more appropriate.

METHODOLOGY

The study was conducted with oversight and assistance from the City of Palo Alto. Without explicit endorsement by the City, employers would have withheld employee addresses from a university researcher.

Palo Alto committed twenty hours of staff time to the project. Part of the research challenge was to decide how to use the City's time most productively and to determine when to call upon the City's influence with these employers. One of the first actions was to get the City's GIS Manager to formally approve the methodology.

Previously, California law required large employers to designate a company "commute coordinator," a person assigned to implement Transportation Demand Reduction strategies. The law has expired and the function is now retained only on a voluntary basis. The role typically reports to one of three departments: human resources, environmental health services, or facilities.

In advance of the launch, a web site was prepared to provide background information on the research methodology. The City launched the study at their quarterly "Mid Peninsula Commute Forum, " a meeting of commute coordinators, ridesharing organizations, and transit agencies. The City sanctioned the study's approach and methodology, and immediately induced three companies to participate.

In parallel with the launch, a contact list for the twenty SRP companies was created. For companies lacking commute coordinators, human resources or facilities personnel were chosen. The City e-mailed a follow-up request to the contacts, shown in Figure 2. In this message, the City emphasized congestion relief, environmental benefits, and employee privacy protection. The City definitively asked for participation and indicated a willingness to follow up with uncooperative companies.

From there, the researcher followed up with an average of 25 contacts per company, via phone, e-mail, and in-person meetings. Considerable time was spent tracking down decision makers and data owners.

Cooperating companies compiled a spreadsheet with address, city, and zip code for each of their employees who worked in SRP. See Table 1.

The researcher then made an appointment for an on-site data processing session. The addresses were copied to the researcher's laptop PC, which is loaded with ArcView 3.2 geographic information systems (GIS) software and a street map database for "geocoding". Geocoding is the process of translating an address into very accurate two-dimensional flattened earth coordinates. These coordinates are then located on a numbered grid of 1/5 mile x 1/5 mile squares (see Figure 3). In Palo Alto, such a grid cell contains about 144 houses. The number of address "hits" within grid cells is tallied. The process of translating addresses into the larger geographic unit of the grid cell is called "aggregation," and was performed under employer supervision. The resultant table of numbered grid cells and the count of addresses within each cell is shown in Table 2. Only non-zero count cells are retained.

Once the addresses were aggregated, the address spreadsheet was deleted from the researcher's laptop PC. Absolutely no addresses left the premises.

In two instances, companies felt comfortable enough to e-mail the address spreadsheet directly to the researcher, provided assurance was made that address data would be immediately processed and deleted.

Methodological Details

The study required 280 hours, with the following breakout: 100 for employer persuasion; 80 for GIS work including geocoding/aggregation, base map creation, and grid database processing; 40 for project management and web site development; 40 to research and refine the methodology; 20 for the City of Palo Alto staff time. In addition to labor, the major expense item consisted of plotting more than 70 3' X 3' high resolution color maps.

Out of twenty companies, thirteen participated. One of the participating companies required "zeroing out the last two address digits," a topic explored in a later section.

For 12 out of 13 of participating companies, the address-list-generating database query took less than 15 minutes to run. For one unfortunate company, data compilation was a painstaking, manual process, and took more than ten hours. Since the study was conducted, a 15-minute method has been discovered at the later company. Two of the declining companies pointed out that, as multinationals, personnel from multiple, independent divisions work in Stanford Research Park (SRP). Some of these divisions are headquartered outside of the U.S. with personnel databases residing in those headquarters; so they claimed address list generation would have been problematic, but this may have been a pleasant excuse to refuse participation.

Table 3 provides the reasons given by the seven declining companies. False reasons may have been provided to ease the uncomfortable task of saying "no." Even if the given objections disappeared at declining companies, a new reason might have arisen to prevent participation. The main reasons given for declining were privacy related. A discussion of privacy issues follows in the section entitled "Data Protection / Privacy."

For a study relying on voluntary corporate participation, the human element is vital. One potential personal style for researchers to use in future studies is "pleasant determination." Company participation may depend solely on the random sequence of contacts made within a company and the current workload of those employees. One overworked employee can decline on behalf of an entire company.

As a template for the initial employer contact, the first three sentences of a researcher's phone conversation with a commute coordinator might identify concepts that are very hard to refuse, "I'm following up on the City of Palo Alto's February 10th e-mail request about the commute study <hard to refuse the government>. Companies A, B, C, and D have already participated <peer pressure>. The goal of the study is to reduce traffic congestion and pollution."

The determined researcher should have a ready answer for most objections. Where "time" or "database" is a problem, remind contacts that other companies have compiled the address spreadsheet in 15 minutes. The researcher should gently make it clear that non-participation leads to escalation, requiring more time than cooperating. For the Palo Alto study, a predetermined escalation procedure was defined where the City offered a face-to-face meeting to discuss any concerns. Where employee data protection is an issue, mentioning the approval of the City's GIS manager and the participation of a leading law firm is appropriate on the phone, but an in-person meeting showing a preliminary map can best explain the reassuring geographical concepts.

The base map was created from Tiger data. Transit route and highway data was provided by the regional transportation agency.

Future Methodological Improvements

The second study (South San Francisco) is underway and should complete in early 2003. A time reduction from 280 hours to 100 hours is expected, for two reasons: reusable work from the first study and time saving refinements. There are five refinements:

First, the companies will be treated as a group. Communication between companies will be encouraged, with three expected benefits: companies will resolve issues together, group interaction will increase comfort with the study, and peer pressure will lead to a higher participation rate. Group data collection will be scheduled for a single time at a central location, reducing data collection time. A government auditor may be provided to guarantee data security. As part of the peer pressure strategy, the largest employers will be persuaded to participate first, to set the tone for the rest of the employment center.

Second, "last two digit zeroing" (see below) will be encouraged. Employers will either bring full addresses to group data collection sessions, or electronically transmit "zeroed" addresses.

Third, the participating government entities will own the study, making it even harder for employers to decline because of lack of time – declining the local government might appear disloyal. In contrast, the first study was only a private study with City of Palo Alto endorsement.

Fourth, a government attorney may issue a formal (and favorable) opinion on data protection issues raised by European Parliament Privacy Directive 95/46/EC, increasing the difficulty to decline based on privacy concerns. The government attorney may be available to address privacy concerns raised by individual companies and to share with all companies any insights that arise during the course of the study.

Fifth, the Congestion Management Agency (CMA) will handle most communication with employers, taking the lead in requesting address data. The CMA, already working with major employers to reduce congestion, enjoys working relationships with many of the companies.

DATA PROTECTION / PRIVACY

Many citizens sense that their personal data (such as name, phone number, address, income, social security number, medical records, and credit history) is being regularly misused or used without their permission. Our everyday barrage of unwanted junk mail and telemarketing calls attests to this fact, but there are many more sinister misuses to contemplate. Proper use of personal data in today's "information overload" society is increasingly on state and federal legislative agendas.

For the Commute Shed Study, no personal data was used. Personal address data was processed, but discarded while under employer supervision. The aggregation of address data to a larger geographic unit, the 1/5 mile by 1/5 mile grid, eliminated the need for address information.

After aggregated grid data was taken away from the employer site, a second level of data protection was applied. Grid data from individual companies was combined together. Furthermore, the names of participating

companies were not revealed in both the resultant map and final GIS data files. Therefore, it is impossible to ascertain which SRP companies may have an employee or employees represented in an individual grid square.

In contrast to the Commute Shed Study, detailed personal data is collected in many surveys, such as the U.S. Census. The challenges of protecting this data from misuse are quite involved.

Because of the absolute lack of personal data, the Commute Shed Study easily complies with U.S. Government Privacy Policy, as embodied by U.S. Census Bureau methodology. Many U.S. multinational corporations with operations in Europe voluntarily apply European Union (E.U.) Privacy Directive 95/46/EC to their U.S. operations. In response to concerns about rampant misuse of personal data, the E.U. requires employers to obtain informed consent from each employee before providing address data to a third party. The E.U. bias is to eliminate casual transfer of such data. Fortunately, the law allows for aggregation of address data to the point where addresses and individuals cannot be identified.

Almost all major U.S. companies have a written privacy policy. Because these policies must by necessity be quickly read and understood, these policies are shorter and simpler than the E.U. Directive and skip complicated topics such as aggregation. Except in a few cases where companies have signed on to the U.S. - EU "Safe Harbor" Agreement, the link between corporate privacy policies and the E.U. Directive is not stated publicly, but exists nevertheless. None of the individual corporate policies "go beyond" the E.U. Directive.

The E.U. law is relatively new and no geographical "aggregation precedent" exists to provide the conclusion that "two level aggregation using 1/5 mile grid and multiple employers" is permitted, but "single level aggregation using a 1/10 mile grid" is prohibited. Instead, E.U. law encourages companies to draw their own conclusions, and allows for interpretation variation even within E.U. member countries. The specific legal test is whether "using the Commute Shed data with all other reasonable means (such as creating a combined query with other public data sets), can we identify an individual that we could not have identified without the Commute Shed data?" (Test provided in April 23, 2002 e-mail correspondence by Diana Alonso Blas, LL.M., European Commission Data Protection Unit, Brussels.) The answer: "No, the Commute Shed data does not reveal much at all."

Four privacy experts provided informal opinions in favor of the methodology: Chris Kelly, Privacy Attorney at Baker & McKenzie and former Chief Information Officer for Excite@Home; Kent Walker, General Counsel at Liberate and former General Counsel for Netscape; Charles Mudd Jr., Chicago attorney and President of Privacy Innovations Inc.; and Bruce Joffe, President of Oakland-based GIS Consultants and Board Member of the Bay Area Mapping Association. Kelly agreed "the grid is not traceable to an individual," and opined "When privacy protection turns into NO DATA, then we have a disaster." An unsuccessful attempt was made to obtain an opinion from the E.U.'s Diana Blas. This attempt resulted in a second attempt at an opinion with a new E.U. contact, Anne-Marije Fontein, International Officer for the Dutch Data Protection Authority. This attempt was also unsuccessful.

While most companies cooperated with the study, six companies declined at least in part because of privacy issues. At four of these companies, attorneys specifically sited the consent rule, but were unaware of the aggregation clarification. The lack of a clear E.U. geographical aggregation precedent made converting these viewpoints doubly hard. A formal opinion from a privacy attorney might have helped convince attorneys who were adverse to accept legal advice from amateurs. The informal opinions listed above were not persuasive.

Especially in the current environment where personal data is frequently misused, declining attorneys were predisposed to shield employee addresses. One particularly sensitive attorney put it this way, "We have to assume that some day this data will get out. Employees will go 'nonlinear' if we give out this data. We are very, very sensitive to this possibility. If we can err on the side of privacy, we do, even if it appears externally to be 'stupid.'" Complicating matters, declining attorneys were hesitant to explore the geographical analysis and had limited training in geography. One attorney advised that the E.U. Directive was so strong that only local, state, or national legislation could compel his company to participate – true for the consent rule but false for the aggregation clarification. However, for a company to VOLUNTARILY provide their address data, the challenge is greater than being on the right side of the law, instead the challenge is often to convince the decision-making attorney to participate.

Methodological Variation: "Zeroing Out" the Last Two Digits

Two companies interpreted the E.U. Directive as forbidding third parties from even temporarily seeing address data, such as during the study's on-site aggregation. An exceptionally aggressive parse of the E.U. Directive results in this conclusion, but the issue more accurately circles back to the assurances by the City of Palo Alto and the

researcher that the address spreadsheet would truly be deleted and the file would not be furtively copied. At some point, a minimal level of trust is required from employers.

An approach was devised to address this flawed interpretation: "zero out" the last two address digits before providing the spreadsheet. 888 Oak Tree Lane, 3333 Pitman Drive, and 1111 5th Avenue change to 800 Oak Tree Lane, 3300 Pitman Drive, and 1100 5th Avenue. This zeroing provides information as to the block that employees live on, without giving up the address itself. Within Palo Alto, blocks have about 36 houses. The zeroing procedure has been used successfully with one company. With zeroing, there is a chance that some modified addresses will aggregate to a grid cell that is adjacent to the correct cell. With popular spreadsheet programs, “macros” can rapidly perform the zeroing. There are a number of variations with zeroing. Address numbers smaller than 100 are currently assigned the value 50.

While an “aggregation precedent” for address data has not been set, this intermediate level of aggregation to about 36 houses most probably also satisfies E.U. law. This creates the opportunity for employers to easily modify their address database into a form they feel comfortable distributing, without the need to run a GIS program on their premises.

Methodological Variation: Zip+ 4

In addition to the variation in the last section, one employer suggested using zip+4 postal codes instead of “zeroed” addresses. Zip+4 and zeroing result in a similar intermediate level of aggregation, but zip+4 adds additional effort and complexity.

In 1983, the Postal Service began using "zip+4," the original 5-digit zip code plus a 4-digit add-on code. To assist efficient mail sorting, the 4-digit code identifies a city block, office building, or individual high-volume location. In residential Palo Alto, each zone covers one side of a block, about 18 houses.

Issues with zip+4 are threefold. First, not all employers maintain zip+4 data on their employees. Second, compared to street map databases, there are fewer zip+4 databases to choose from. Third, the U.S. Postal Service changes zip+4 codes on a regular basis, and does not guarantee accuracy for any period of time.

U.S. Privacy Policy

The U.S. does not have a single, consolidated privacy policy. Privacy laws are passed to address only narrow issues and policy-making is distributed through multiple government agencies. For the purposes of the commute shed, Census Bureau privacy policy serves as a proxy for governmental policy. A Census example of how data is “anonymized” is provided, followed by an explanation of Census law. Finally, a commonsense personal medical data example provides further explanation of the issues involved.

U.S. Census Example

Some revealing Census data is tabulated in the Public Use MicroSample (PUMS). PUMS data is consciously cleaned of information that could allow identification of an individual. "In addition, detailed personal data that might allow the identification of a given person because of his or her unusual occupation, income, or family situation, has been adjusted to reflect somewhat less exotic characteristics. The effect of these adjustments is that it would be impossible to find Bill Gates in the Washington State PUMS, even if he filled out the 5 percent Census sample instrument-because some of his distinguishing income and asset features would have been 'topcoded' or replaced by other values that captured the spirit, but not the detail of high income or wealth." (5)

U.S. Census Law

Section 9 of the Census Bureau’s Title 13, “Protection of Confidential Information,” prohibits non-statistical uses, release of individually identifiable data, and examination of individual surveys by non-Census personnel. The prohibitions follow:

> (1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or
> (2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or
> (3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports. (6)

Section 214 of Census Tile 13, “Wrongful Disclosure of Information,” provides for fines of up to $5,000 and jail terms of up to 5 years for violators of Section 9.

Common Sense U.S. Aggregation Example

A medical study tracked "at risk" infants whose mothers had not partaken of Washington State "First Steps" preventative health services for expectant women:

"Each of the data points contained a wealth of information about the mother and each birth, but such information is confidential; <the researcher> could not simply publish a map of Yakima County's Medicaid recipients. As GIS becomes a more commonly used tool in government, especially in social services, GIS managers are having to resolve this conflict between confidentiality and data that can give researchers valuable information about social problems.” (7)

The researcher aggregated the information into census block groups, containing 250 to 550 addresses, “an area large enough to remove the identification of an individual address, but small enough to connect a trend with an identifiable area.” (7) Unlike the Commute Shed Study, this medical data set contains detailed and potentially damaging personal information, which may also be of use in identifying individuals.

European Union Privacy

The European Parliament Directive on Data Protection, 95/46/EC, was passed in October of 1995 and went into effect in October of 1998. Three important components of the law are: informed consent, anonymization, and public interest. Swire, Fromholz, Cullen, and Salbu provide detailed analyses of the Directive. (8, 9, 10, 11)

Recital 28 requires employers to obtain consent from employees before using personal data for a different purpose than it was originally collected. An employee provides personal data upon being hired. The employer provides a description of data uses at that time. Under this recital, unless the employer explicitly includes something like “address data will be used for transportation planning studies,” the employer must obtain a separate consent for such use.

Recital 26 introduces the concept of anonymization, “in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible … using all means that are reasonably at their disposal.” (12) This anonymization overcomes the consent provision.

Recital 30 further weakens the consent provision for commute shed studies by allowing collection of personal data without consent for tasks “in the public interest … , provided that the interest or the rights and freedoms of the data subject are not overriding.” (12)

The Directive is an imprecise law that does not provide clear priority between competing data protection concepts. (13) With the wealth of public databases available to combine with survey data, the distinction between anonymous and identifiable is not simple. Very small amounts of aggregation have been rejected. In one instance, the use of a license plate number was deemed intrusive because it identified a small set of possible drivers of a particular vehicle. (14)

U.S. – European Union Safe Harbor Agreement

The European Union (E.U.) Directive prohibits transfer of sensitive personal information from the E.U. to non-E.U. countries where countries do not provide an “adequate” level of data protection. In Year 2000, to avoid interruption of sensitive data transfers, the U.S. and E.U. negotiated the Safe Harbor Agreement. U.S. multinational companies were allowed to transfer data provided they agree to follow and self-regulate the E.U. Directive, with some Federal Trade Commission oversight. (15,16) U.S companies certify themselves on a Department of Commerce web site so that E.U. companies can verify permission to transmit information. Of the declining SRP companies, half were certified on the Department of Commerce web site.

CONCLUSIONS

The 280 hours spent on the first study were longer than expected. With the planned improvements for the second study, the methodology will fulfill the promise of rapid, precise origination data collection, and will become more attractive for wide deployment. Should the process to be automated to the point where employers can self-aggregate, wider deployment is almost assured.

E.U. privacy law is newly effective as of 1998 and data protection legislative activity is increasing in the U.S. Precedents should be developed to define the "breaking point" between permissible and prohibited aggregation.

ACKNOWLEDGEMENTS

Thanks to U.C. Berkeley City and Regional Planning Department (Professors Robert Cervero and John Landis), GreenInfo Network (Lynn Frederico, Brian Cohen), City of Palo Alto (Joe Kott, Amanda Jones, Dave Matson), Santa Clara County Valley Transportation Authority (Chris Augenstein), the nine county Bay Area Metropolitan Transit Commission (Mike Skowroneck), Electric Power Research Institute (Jim Galanis), Anthony-Maymudes Foundation, and Pinnacle Systems.

REFERENCES

1. "ABAG Projections 2000", Association of Bay Area Governments, 101 8th Street, Oakland, CA 94607, 510/464-7900.
2. "Fall 1999 Commute Service Study", Santa Clara Valley Transportation Authority, 3331 N. First Street, San Jose, CA, 95134, (408) 321-2300.
3. "World Class Transit for the Bay Area", Bay Area Transportation & Land Use Coalition, 414 13th Street, 5th Floor, Oakland, CA 94612 , (510) 740-3150. January 13, 2000. Chapter 3: Creating a Regional Express Bus Web. http://www.transcoalition.org/wct/WCT_nopictures.pdf.
4. "On-site ArcView 3.2 Procedure for Commute Shed Study", http://www.cities21.org/ArcViewProc4study.pdf .
5. "Environmental Justice in the Oahu Metropolitan Planning Organization's Planning Process Report," section entitled: "Census Confidentiality Limitations on Disclosure of Income Data," Oahu Metropolitan Planning Organization, 707 Richards Street, Suite 200, Honolulu, 96813, (808) 587-2015. http://www.eng.hawaii.edu/~csp/OMPO/T6EJ/Final2001/Final_Appx_B.PDF .
6. U.S. Census Bureau, Title 13, Section 9, "Protection of Confidential Information," U.S. Census Bureau, Washington DC 20233, 301-763-4636. http://www.census.gov/main/www/policies.html#privacy
7. "GIS in Public Policy"; Richard Greene; ESRI - 380 New York Street, Redlands, CA 92373; 2000; ISBN 1-879102-66-8. Pages 30-31.
8. "European Parliament Directive on Data Protection, 95/46/EC," European Commission Data Protection Unit, Brussels, October 1995. http://europa.eu.int/comm/internal_market/en/dataprot/.
9. "Application of a Methodology Designed to Asses the Adequacy of the Level of Protection of Individuals with Regard to Processing Personal Data: Test of the Method on Several Categories of Transfer," by Charles D. Raab, Colin J. Bennett, Robert M. Gellman, Nigel Waters, University of Edinburgh, September 1998. Published by: European Commission Data Protection Unit, Brussels. http://europa.eu.int/comm/internal_market/en/dataprot/studies/adequat.pdf
10. "Data Protection Law and On-line Services: Regulatory Responses," by Joel R. Reidenberg, Fordham University School of Law and Paul M. Schwartz, Brooklyn Law School. Published by: European Commission Data Protection Unit, Brussels. http://europa.eu.int/comm/internal_market/en/dataprot/studies/regul.pdf:
11. "None of Your Business : World Data Flows, Electronic Commerce, and the European Privacy Directive," by Peter P. Swire and Robert E. Litan. 1998. Published by: Brookings Institution Press, Washington D.C.
12. "The European Union Data Privacy Directive," by Julia M. Fromholz in Berkeley Technology Law Journal, Annual Review of Law and Technology: VI. Foreign and International Law. 2000, volume 15.
13. "A business guide to changes in European data protection legislation", by Cullen International. 1999. Published by: Kluwer Law International, Cambridge.
14. "William Davidson Working Paper Number 418: The European Union Data Privacy Directive and International Relations," by Steven R. Salbu. December 2001. Published by Davidson Institute, 724 East University Avenue, Wyly Hall, First Floor, Ann Arbor, Michigan 48109.
15. "U.S.-EU 'Safe Harbor' Data Privacy Arrangement," Edited by Sean D. Murphy in American Journal of International Law: Contemporary Practice of the United States Relating to International Law. Volume 95, Issue 1, January 2001, pgs 156-159.
16. "Safe Harbor and the European Union's Directive on Data Protection," by Jordan M. Blanke in Albany Law Journal of Science & Technology, Volume 57, Year 2000.

LIST OF TABLES AND FIGURES

TABLE 1 Address Spreadsheet with Imaginary Addresses

Address	City	Zip Code
555 Hans Ave.	Mountain View	94040
333 Crestview Ave.	Mountain View	94040
888 Oak Tree Lane	Mountain View	94040
3333 Hastings Lane	Mountain View	94040

TABLE 2 Resultant Grid Cell Counts

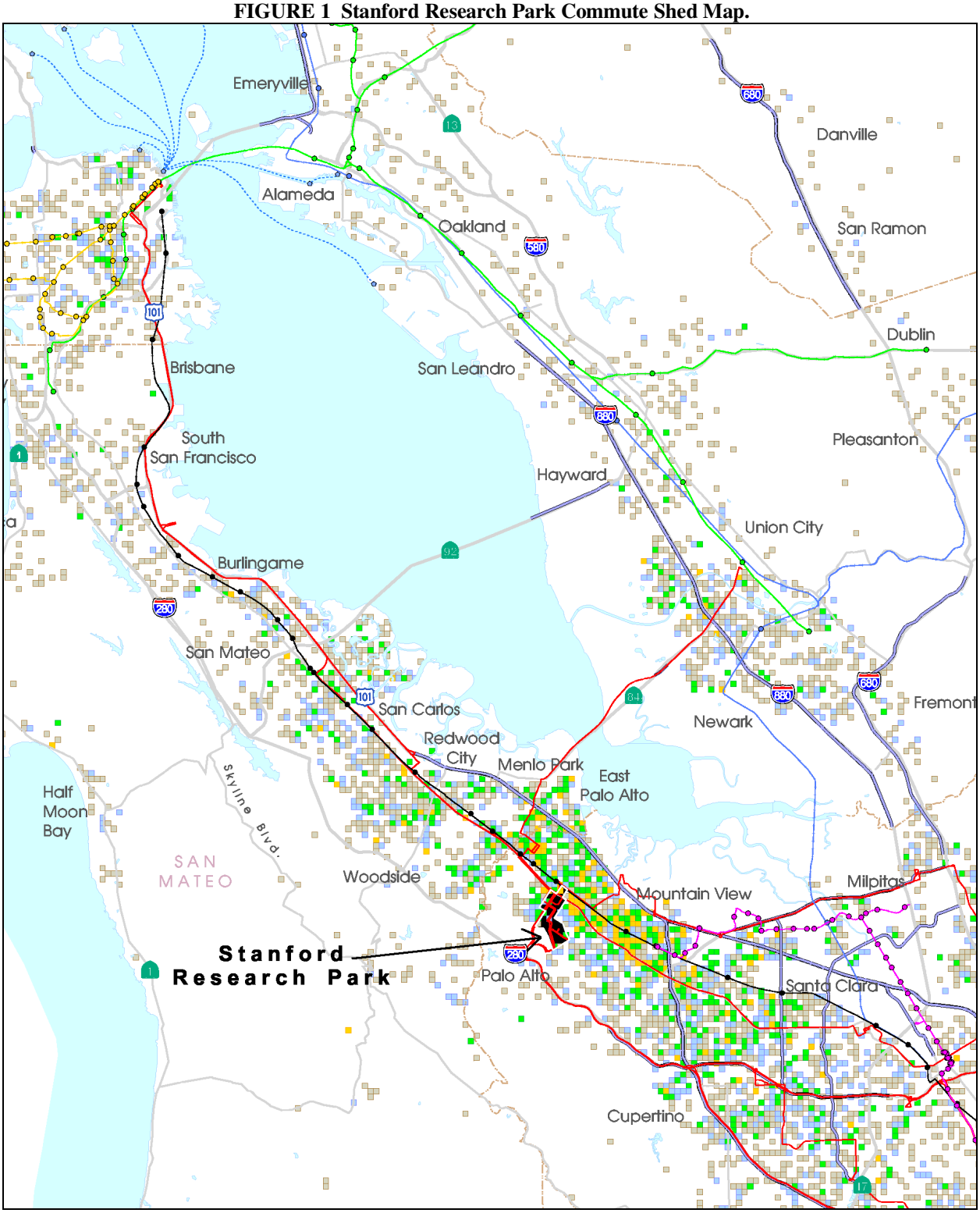
Grid Number	Count
12023	1
62665	2
113455	1
124293	1

TABLE 3 Reasons Given by Declining Companies

Company	E.U. Privacy	Genl. Privacy	Time	Database
1	20%		80%	
2		100%		
3				100%
4		100%		
5	80%			20%
6	33%		66%	
7	80%			20%

"E.U. Privacy" means the reason given was the company follows a data protection policy similar to European Parliament Privacy Directive 95/46/EC. "Genl. Privacy" means the reason given was "our policy is not to give out that information." After this response, the contacts indicated that further contact was unwelcome. As this was a voluntary process, the contacts were not obligated to expand upon their refusal. "Time" means that the reason given was the company did not have enough time to participate in the study. "Database" means that the company felt that extracting the addresses from company databases was going to be objectionably hard. For the company given a 100% score for "database," the contact admitted that lack of training on the new HR database had left them unable to query for addresses. For mixed scores, such as 80% "time" and 20% "E.U. Privacy", the company declined because of a primary issue, but indicated that a secondary issue also loomed as an obstacle.

FIGURE 1 Stanford Research Park Commute Shed Map.



Legend: brown 1/5 mile grid cells contain 1 person, blue cells contain 2 people, green cells contain 3 or 4 people, yellow cells contain 5 to 18 people. Red lines are selected bus routes. The black rail line is Caltrain commuter rail. Thick blue-purple highways designate HOV lanes. Please see http://www.cities21.org/final_map.htm for larger versions of the map. The actual printed map is 36" x 36" at 600 DPI and covers roughly four times more area.

FIGURE 2 Letter to SRP commute coordinators.

Dear Commute Coordinator,

The City of Palo Alto has approved and endorsed a "fine-grained commute origination" research study for the approximately 20,000 employees in Stanford Research Park. The study will **improve transportation planning** within SRP - assisting transit agencies, ridesharing agencies, and the City of Palo Alto. In order to reduce traffic congestion, polluted auto emissions, and oily highway runoff into the bay, the City of Palo Alto STRONGLY urges SRP companies to contribute to the study in a very timely manner.

This study builds upon the Fall '99 VTA "zip code" Commute Service Study and maps home originations to a much finer scale. The scale is coarse enough to **protect the privacy of employees**, but fine enough for micro analysis (answering "how many people are within 1,500 feet of a bus stop?" for instance). This origination study will help achieve the level of trip reductions demanded under the Stanford General Use Permit.

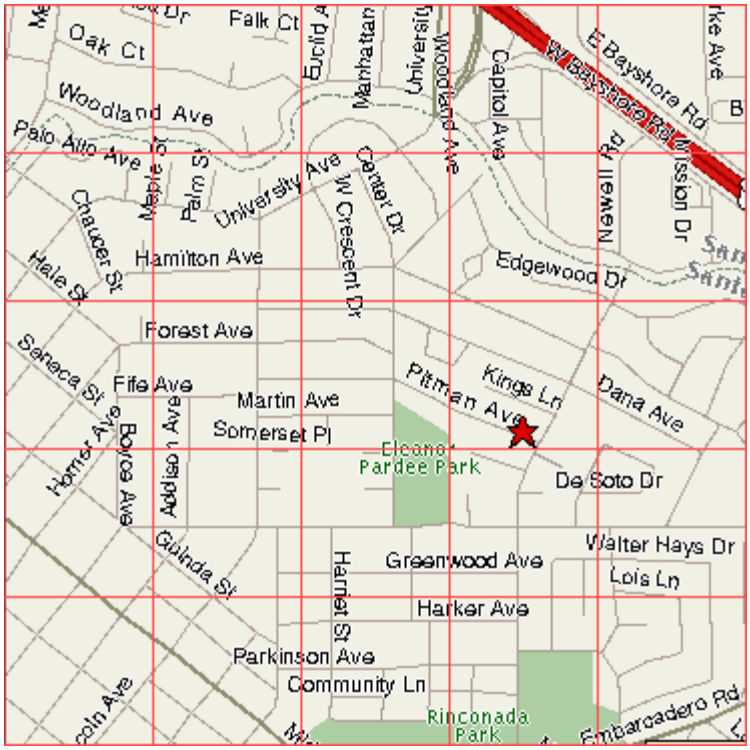
The study is being conducted by Steve Raney, a Cities21 and U.C. Berkeley researcher. Please contact Steve to schedule a short data collection session at your facility. Company IS personnel are encouraged to observe the procedure to increase your assurance of security. The study methodology is sound, simple, and will protect employee address information. We have chosen to work with Steve for his high integrity. Steve's contact info is: Steve Raney, steve_raney@cities21.org, (650) XXX-XXXX, XXXX Pitman Ave., Palo Alto, CA 94301. For any cases where companies are unwilling to cooperate, we will personally attempt to address any concerns your company may have.

For further details on the study methodology, please see http://www.cities21.org/srpOriginationStudy.htm.

Cordially yours,

Joe Kott, Chief Transportation Officer
Amanda Jones, Transportation Systems Management Coordinator
Dave Matson, Geographic Information Systems Manager
City of Palo Alto
250 Hamilton Ave.
Palo Alto, CA 94301

FIGURE 3 1/5 mile x 1/5 mile grid.



Each cell contains about 144 houses.