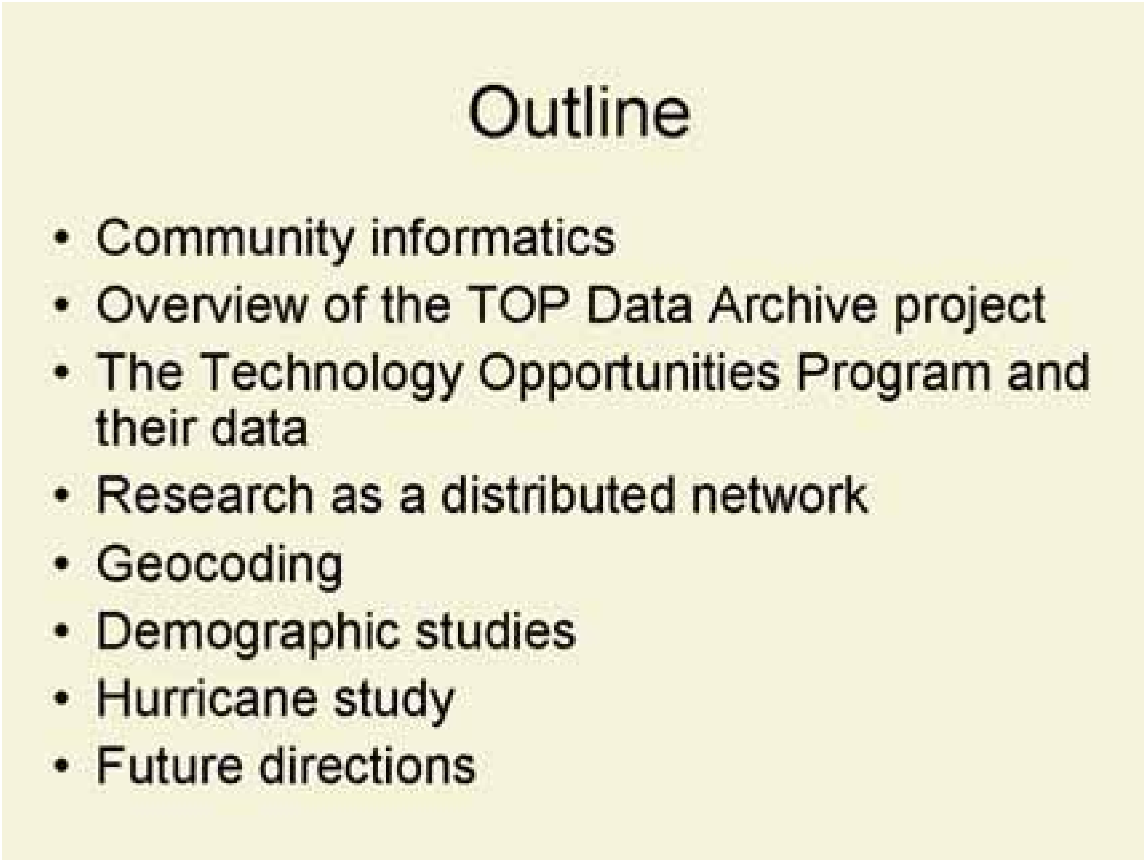
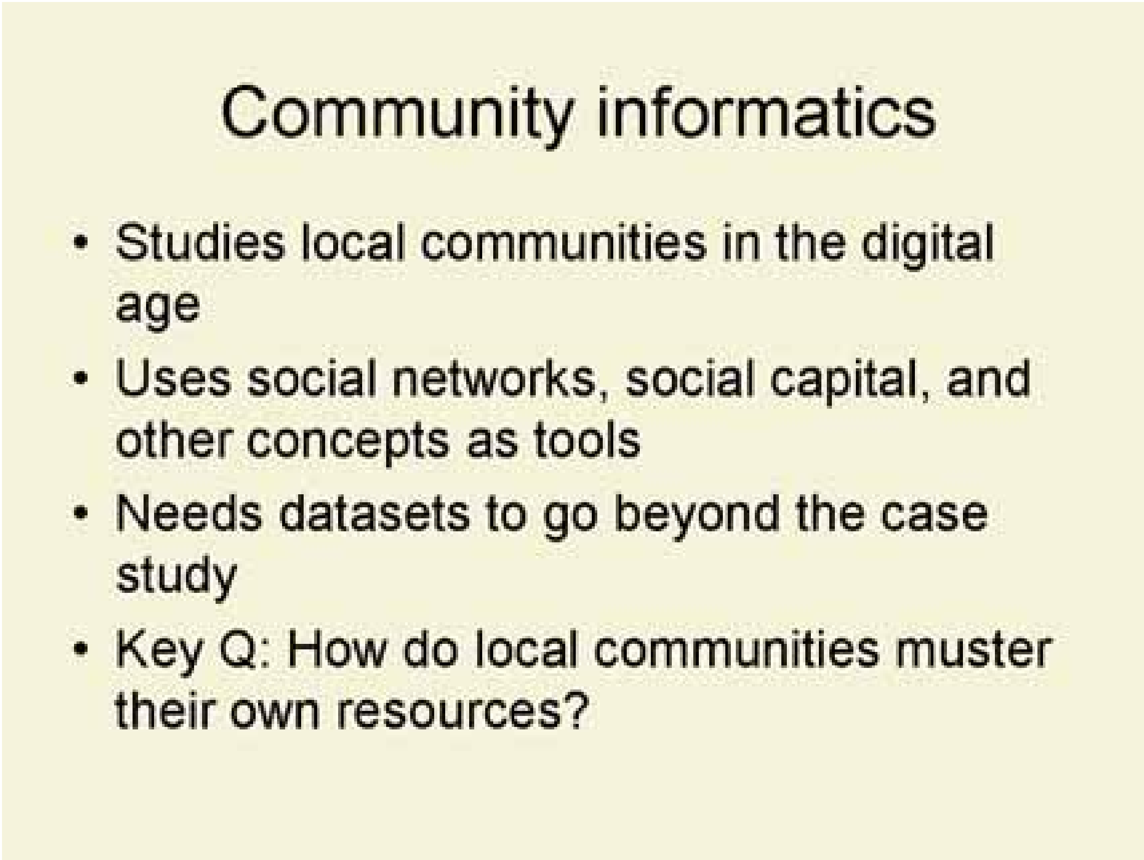
Community informatics and GIS: Exploring a new research collaboration
By Kate Williams (katewill@umich.edu) and Tracey Hughes (hughests@umich.edu)
August 6, 2006

The form of this paper is an experiment in a new format for presenting research. It combines our written text with the slides that will accompany our presentation at the ESRI Education Users Conference in San Diego, California in August 2006. We hope this format makes our story more accessible and more engaging, and we welcome your feedback.

Also by way of introduction, since this paper will be presented in a panel session titled "College/University Collaborations with Local Organizations," it's important to understand from the start that this collaboration is rather unique and large, thanks in part to the power of GIS and in part to the size of the United States. The primary partnerships are a federal agency and the 600-plus local projects they have funded.



This paper will introduce an emerging research area—community informatics—and demonstrate how GIS work is playing a leading role in helping this field grow. The source of our data is a federal agency, the Technology Opportunities Program within the Department of Commerce. Our research team is rather uniquely distributed, and we will explain that. Then we'll review just how we did our geocoding, which involved some up-to-date tricks of the trade. Next, we'll present just two ways that we are using the results of geocoding, of spatializing our data. One is to focus in on the effect of the terrible 2005 hurricane season, and the other is to use demographic data to sift through our entire dataset. In closing, we will present some next steps in this research project.

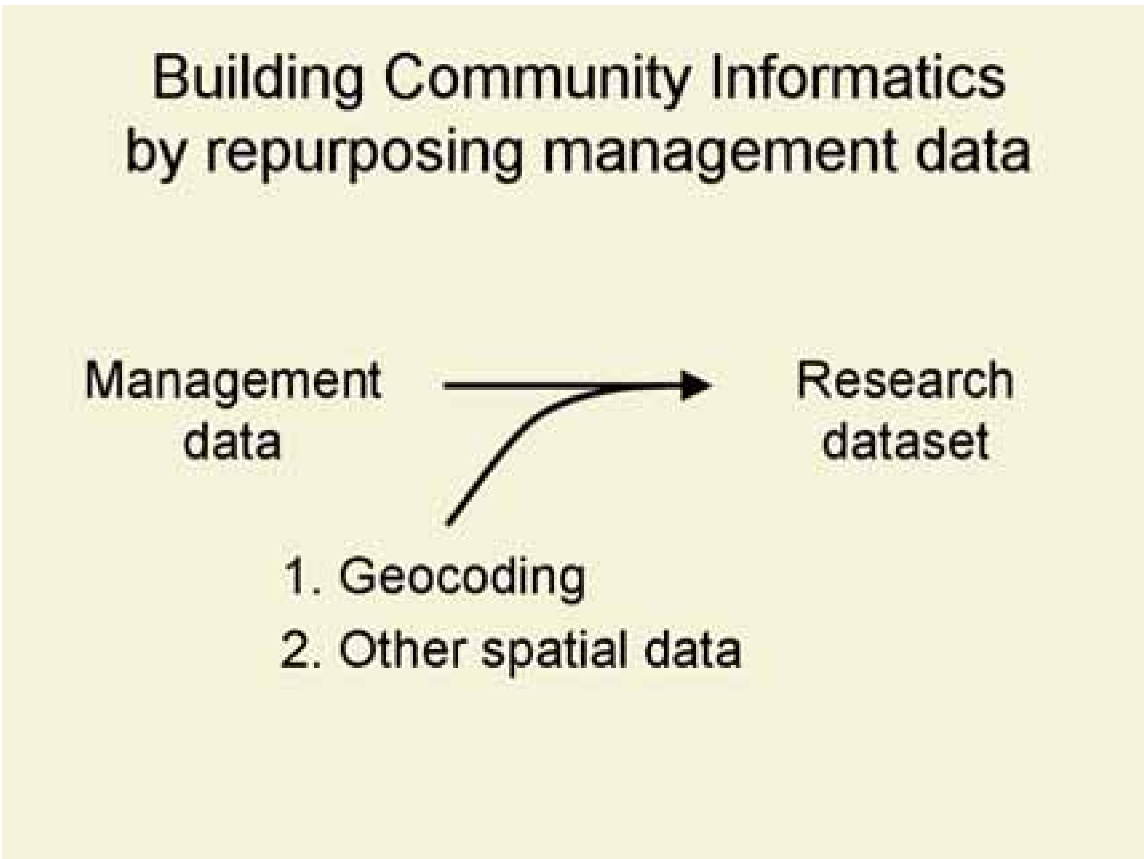


We are operating within the emerging field of community informatics. Community informatics studies local communities in the digital age. This involves examining the dynamic interaction between the historical—geographic communities where we live—and the transformative—the information technology revolution. Social capital and social networks are particularly salient conceptual tools here, because these concepts get at the internal structure of communities. Communities are made up of people who are connected to each other in various ways. These connections, taken all together, form a network. And the information and activities that we share with each other make up a community's social capital—what we can get accomplished by working with and helping each other. So the concepts of social networks and social capital bring to light the often invisible relationships that make up human communities.

The field of community informatics was launched very much by social action and a research agenda, both focusing on the digital divide during the 1990s. The digital divide is the social gap between users and non-users of information technology; it has since evolved to something much more complex, which we call digital inequality or inequalities. Back in the 1990s though, as projects to bridge this divide unfolded, researchers participated, and they analyzed what happened. So various localities were represented in the literature as case studies.

Case studies have the advantage of specificity and nuance, but cannot easily establish generalizability. For this you need larger datasets. Our project demonstrates that GIS can help us build and use these datasets.

The key question in community informatics, with communities generally experiencing shrinking local budgets and shrinking support from federal, state, and private entities, is the question of sustainability: Where and how are local communities able to muster their own resources to participate in the digital age, however they define that? Can pulling on your own bootstraps help you move into the information age?

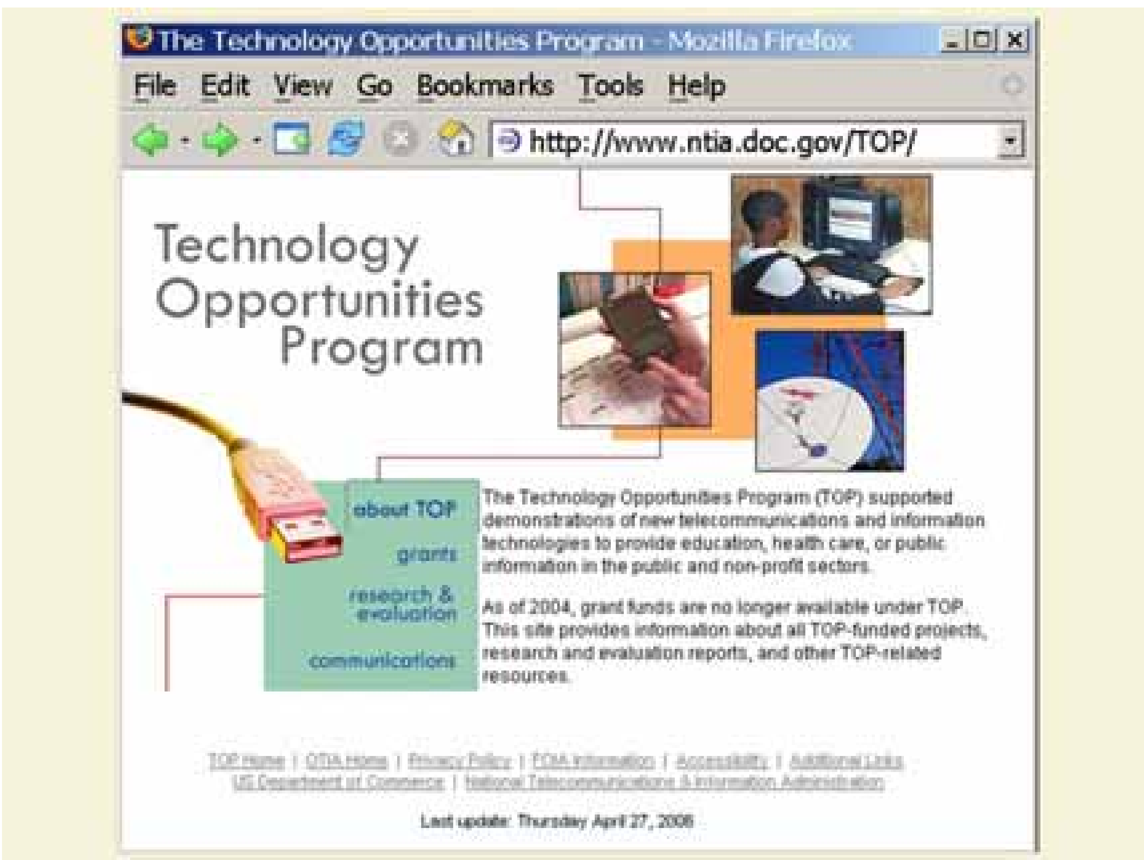


Our approach in this project has been to acknowledge and build on the experience of one particular agency that has been working in this area. In partnership with them, we are repurposing their data—which they collected in order to manage their work—into a research dataset. They have been quite happy to participate in this, because their mission always included demonstrating the value of their local projects, promoting and explaining their projects.

GIS enabled this repurposing process. With GIS, we were able to enrich the dataset by bringing in other data. We brought in two other datasets which we will discuss below. One of these was from FEMA, the Federal Emergency Management Agency, where we extracted data on hurricane paths. The other was the U.S. Census Bureau, where we obtained demographic data from the 2000 Census and the Current Population Study.

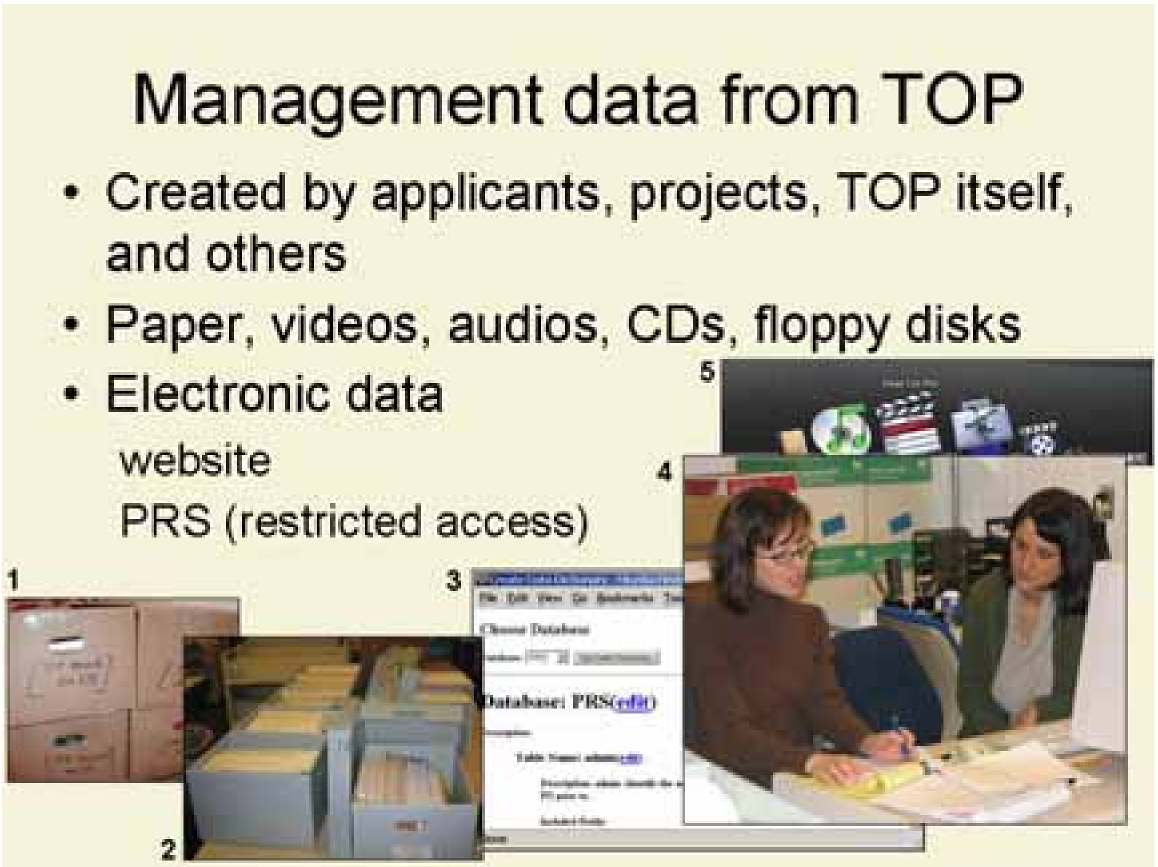
Overall, our goals with this dataset are to use it for our own research and to make it available to others via a website, an online data repository, and a physical archive. So we are providing the first shared dataset for the field of community informatics.

Our project is also a demonstration. We want to demonstrate how useful this repurposing can be for research and knowledge. In the digital age, the daily process of management in every sector of our society—business, education, government, community—creates digital records. Because these records can be easy, even free, to copy, researchers can seek them out and negotiate their swift re-use, thus moving science along that much quicker.



The agency that is our partner is within the U.S. Department of Commerce. It's called the Technology Opportunities Program, or TOP for short. It began in 1994 as the Telecommunications and Information Infrastructure Assistance Program (TIAAP). This was a time when the information superhighway was a new idea. In 1993 the World Wide Web was first made publicly available. Soon after, describing a mobile computer lab that had been built inside a truck, our then-Vice President made this comment in a May 29

1996 speech: “It’s rolling into communities, connecting schools in our poorest neighborhoods and paving over the digital divide.” Mobile computer labs were just one example of the wide variety of local projects that TOP funded.



The data that TOP had collected since 1994 originated from different entities: groups that applied to TOP for funds, groups that TOP did fund, TOP itself, agencies related to TOP, and even the mass media. The collection includes paper—such as grant applications—as well as videos, audios—such as a recording of an entire TOP conference—and CDs and floppies that were produced by TOP projects. One particular electronic dataset begins in 1998, when TOP implemented a web-based system for project reports and staff feedback. This is the Project Reporting System, or PRS. Initial project reports were on the TOP website from the start, as part of their mission to promote and educate people about their work.

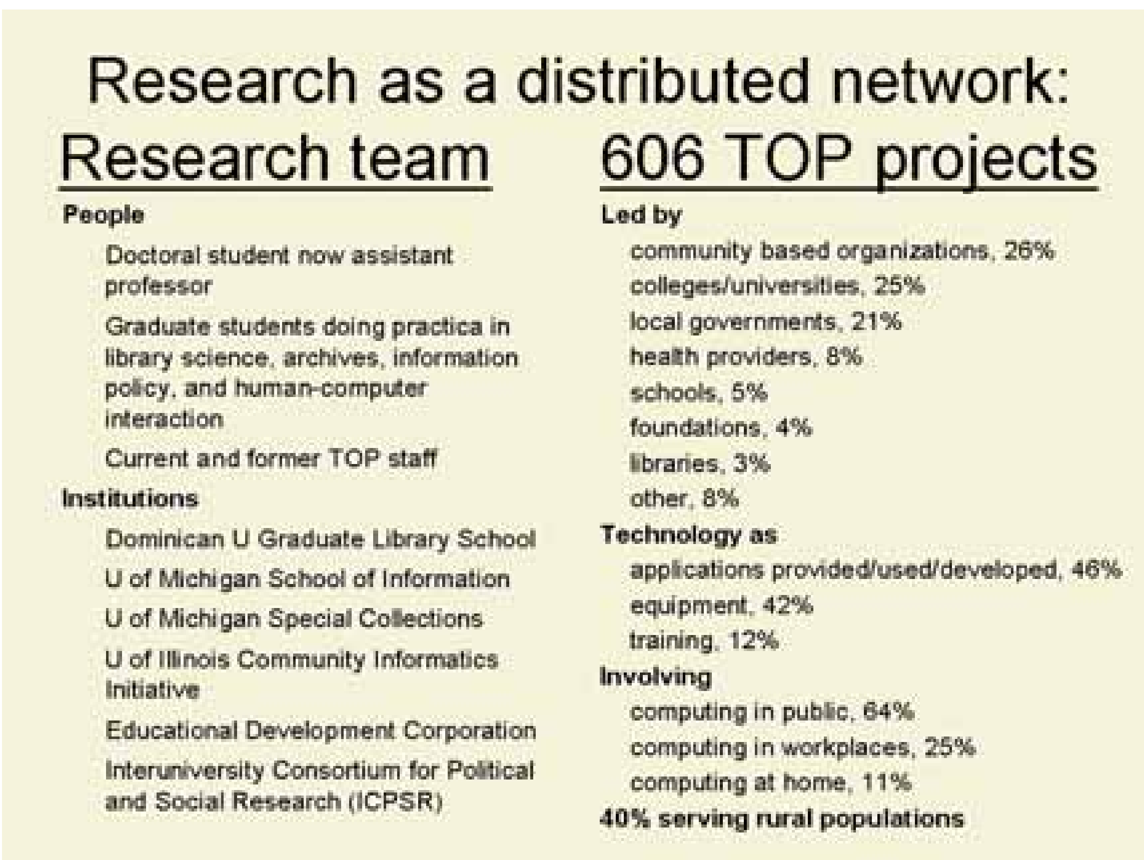
So not only did TOP staff assemble the physical materials and send them to the University of Michigan library (image number 1 in the above slide), they also helped extract this electronic material.

This extraction was possible because one of the archive team members, a master’s student in information policy, took his programming skills to a summer internship at the

Library of Congress, and we arranged for him to spend some of his summer time in the TOP office. This student worked with the very same TOP staffmember who had implemented the PRS in 1998. Besides extracting the data, they created a data dictionary for future researchers (image 3), and even unearthed paper documents that told the story of the database's design origins.

It is important to note here that as we move all the TOP materials into the TOP Data Archive for use by others, use of the PRS database will be restricted to scholars and institutions who commit to protecting the privacy of those in the database. This is because the local and federal individuals spoke quite honestly in the database as their work was unfolding and their confidentiality must be respected.

Another masters' student took on the task to reorganize the physical material in the boxes, according to archival standards. She followed established practices regarding organizing and describing the contents of the collection we are building. This involved putting every item into acid-free file folders and boxes in a particular logical, historical order, and creating what is called a finding aid, with input from one of the university's archivists (image 4). Three other students worked specifically with the video collection and made headway towards creating a digital video collection that can be perused online. Image 5, the Macintosh screen capture, references our software tool for that project, FinalCutPro.

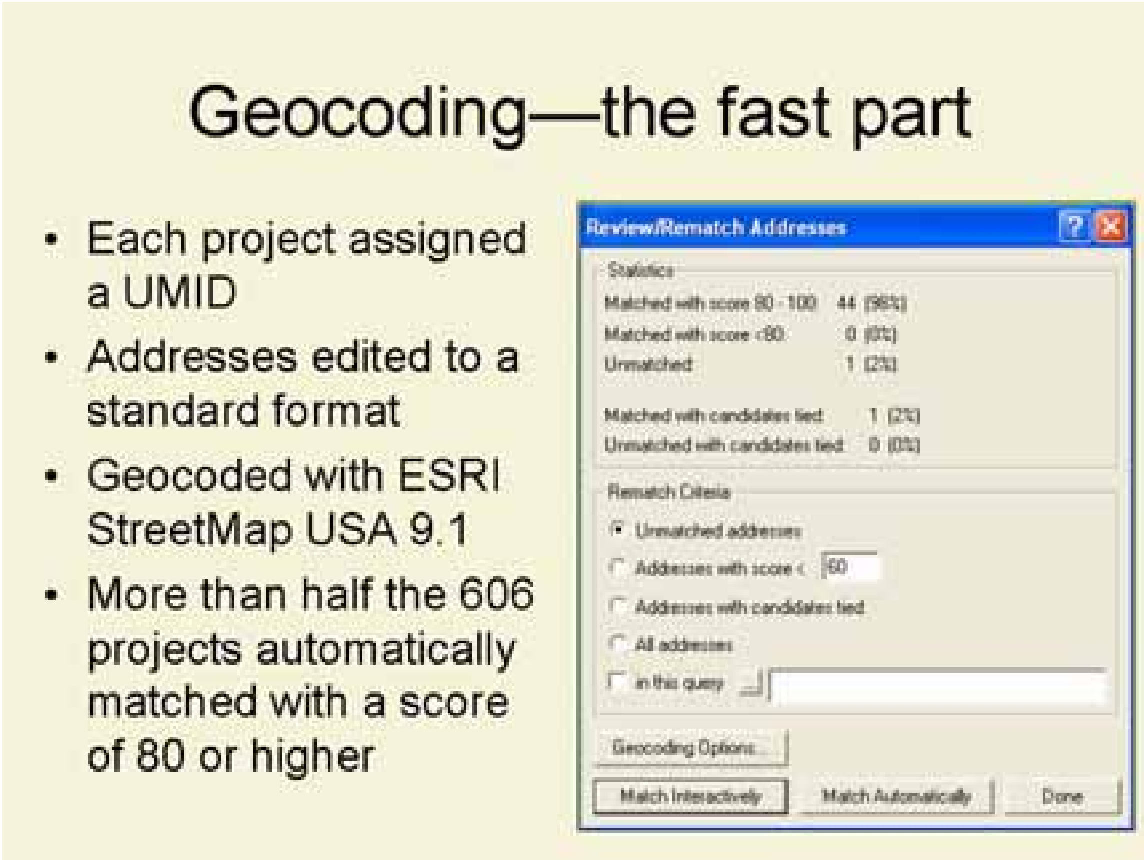


At this point it is meaningful to describe the distributed nature of this work. Like a lot of projects today, it are quite distributed. Six different institutions are carrying out aspects of the work. These are listed above on the left. Also listed on the left are the core individuals, those who have carried out the bulk of the work. One is the doctoral student who proposed the project to TOP, who is now a faculty member at Dominican as well as holding a courtesy appointment at Michigan's School of Information in order to facilitate this project. More than a dozen master's students have also worked on the project as part of the required practicum, or field work, included in their degrees.

In addition, current and former TOP staff have advised, helped move materials to Michigan, and generously told their stories into digital recorders, creating an oral history as part of the overall archive. The spirit of TOP was not so much a routine federal bureaucracy as a reform movement to help level the digital playing field, help ensure "democracy in the information age"—which was even the title of a book by one of TOP's directors. So their orientation was very much to support and join this archive-plus-research project.

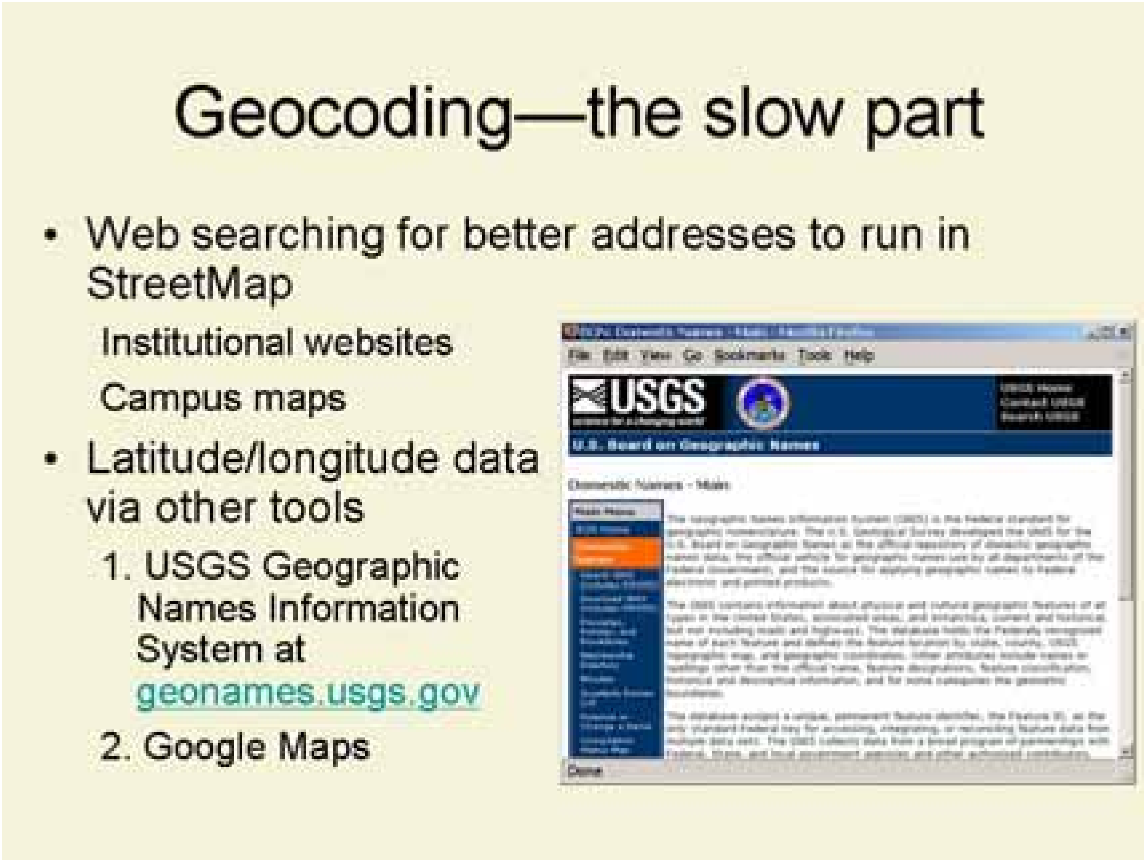
The other large segment of people in this project are the staffs of the 606 projects. They very much share the spirit of mission, of community transformation, that the TOP staff have expressed. Work with them has just begun, but they are likewise interested in telling their experiences and even in contributing their organizational records to the archive.

The statistics on the right side of the slide above give a picture of the TOP projects. In each case the TOP project was a partnership of local organizations funded by TOP for roughly three years. In this slide the lead agencies are identified. Three quarters of them are community based organizations (for instance local social service agencies or local heritage groups), institutions of higher education, or units of local government, which includes tribal governments. They either provided equipment or developed and implemented new applications, with a small percentage focused primarily on training. Two thirds of them enabled computing in public places, such as libraries, community centers, schools, and so on, with the rest concerned with computers in workplaces or in homes. And 40% of the projects have served rural populations, which experience a particular form of digital inequality due to a combination of geographic remoteness and poverty.

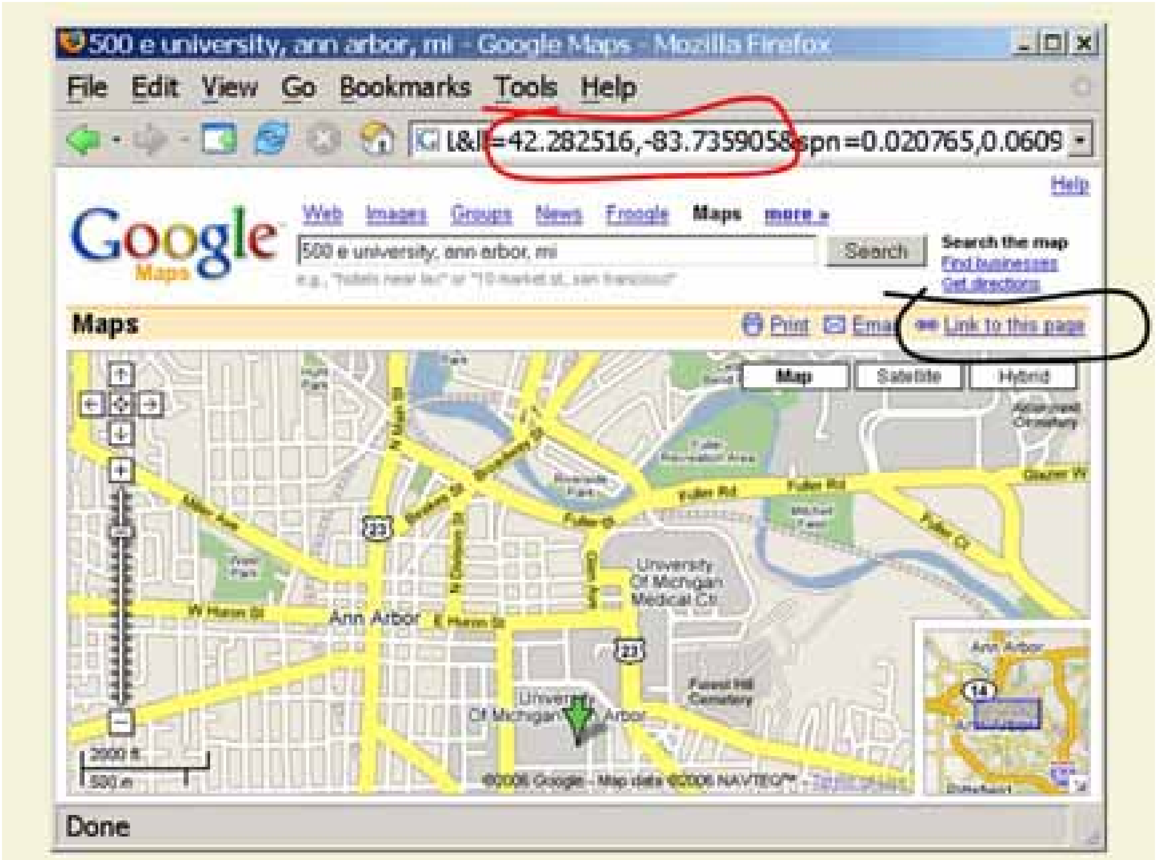


Geocoding of the project was conducted in two phases that can be called the fast part and the slow part, more automated and more handcrafted. The fast part was done using ESRI StreetMap USA 9.1. Slightly more than one-half of all TOP sites were automatically 'matched,' with a minimum match score of 80.

Each project was also assigned a unique number from 1 to 606—a U of M ID number—so that researchers could easily drawn upon and interconnect information from the various parts of the data archive—finding aid, website, paper documents, video, audio, CDs, floppy disks, GIS files, and PRS database.

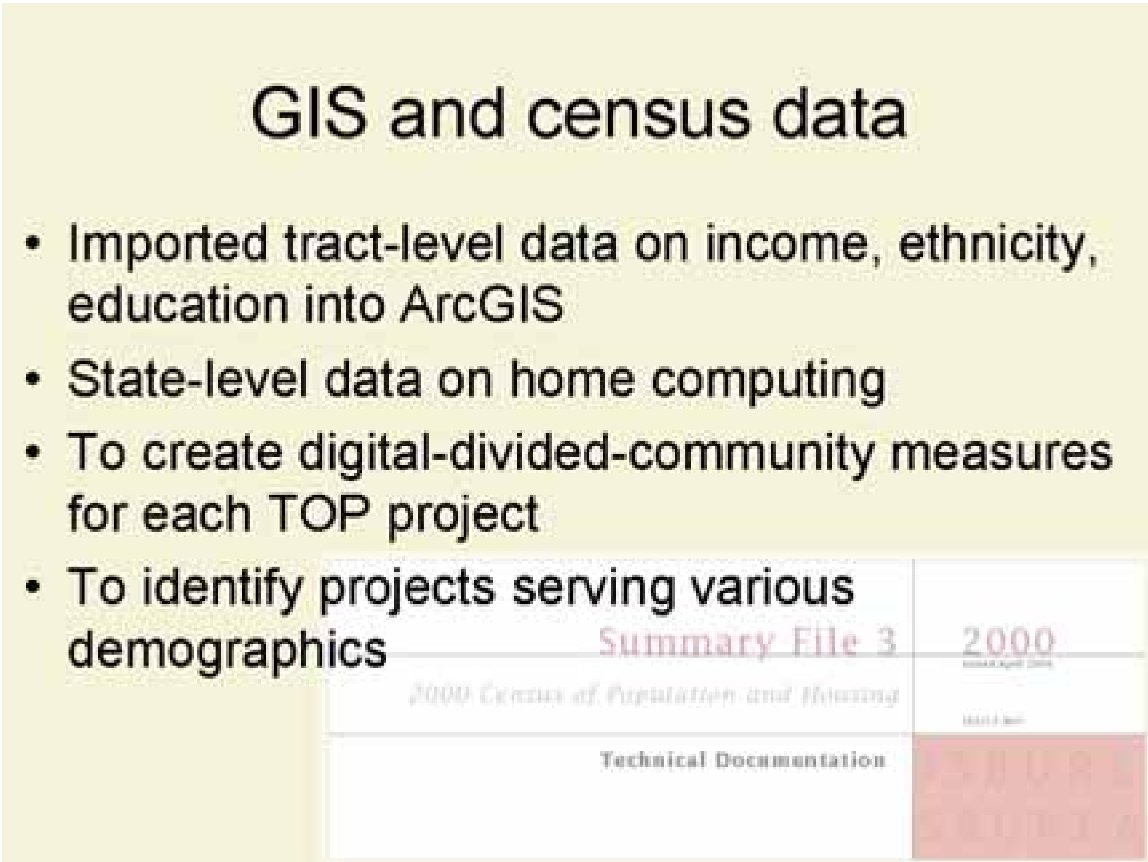


The second part of the geocoding was slow because it was conducted manually. Much of this stage involved extensive web searching to obtain current address information. Institutional or educational facility websites proved very beneficial as did campus maps and directions to a site. After the new address (or in many cases street intersections) was obtained, that data was input into ArcMap and interactively geocoded with StreetMap USA. In the cases where no address information matched with StreetMap, latitude and longitude data was collected from either the USGS Geographic Names Information System (GNIS), located at http://geonames.usgs.gov/, or from the Google Maps "link to this page" functionality at http://www.maps.google.com.



The slide above demonstrates how to use Google Maps to obtain latitude and longitude. When you click on a Google map to create a center point, then click on "Link to this page" at top left, the URL changes to include the latitude and longitude of the center point you clicked on. You can then copy and paste those numbers into the GIS file.

In addition to obtaining latitude and longitude information for each of the 606 TOP sites, address updates were made to a master database of TOP contacts, detailed notes were recorded to clearly identify the source of each address and the match score for that address. Capturing and managing this metadata is critical to the project for two reasons: 1) in sharing this data with the research community, proper documentation for those future endeavors is critical; and 2) with the project originating at an information school, it could only present top quality metadata!



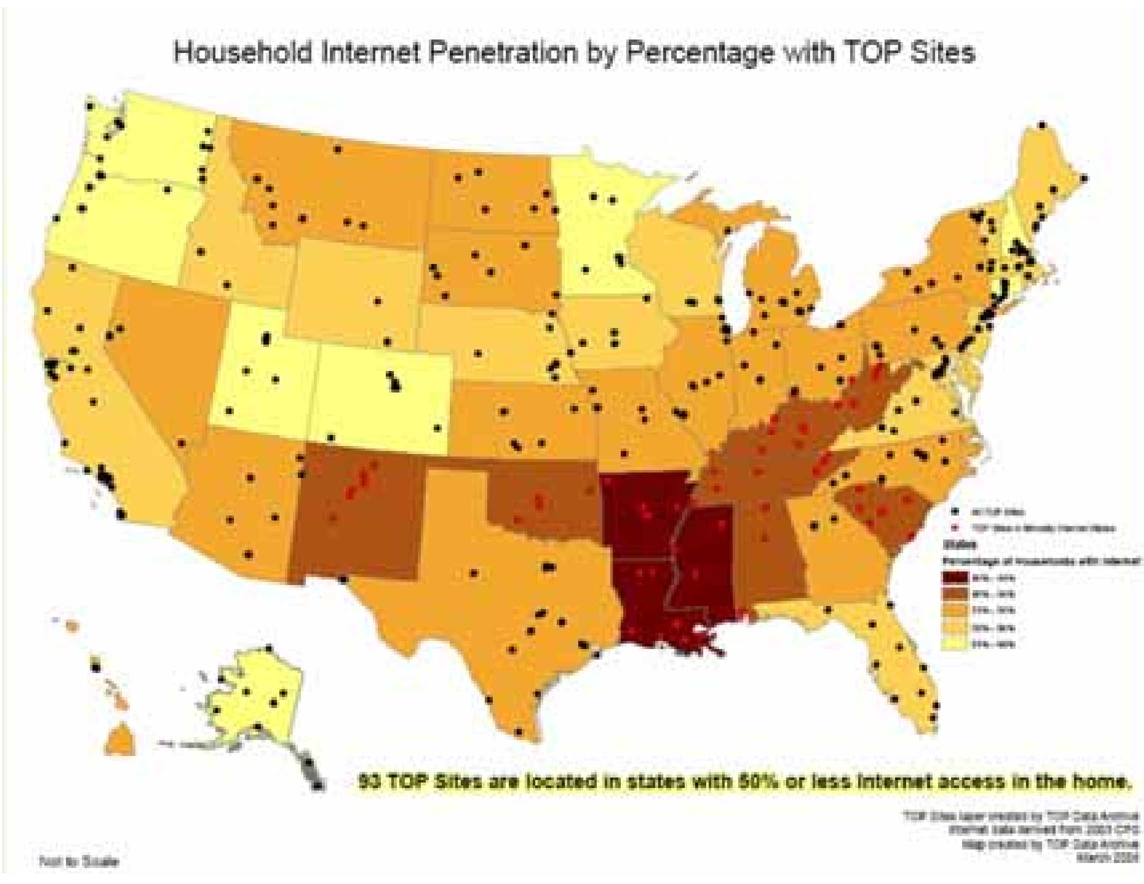
The Census Bureau was our source for demographic data. The idea was to obtain indicators that could be used to indicate the "digital dividedness" of the people living in communities around the TOP projects. Studies done from 1994 to 2001 by the U.S. Census have established that that income, education, ethnicity, and age are associated with different rates of computer and Internet use. Younger people, higher incomes, higher educational attainment, and Asian and white ethnicities are associated with higher rates of access and use, while Blacks and Latinos typically have lower rates. This is all within a context of a broad shift where everyone is moving (at different speeds) towards higher rates of use.

A first step in locating census tract level data on the demographic variables that are associated with digital inequality was to explore the Geolytics product—a data compilation that utilizes U.S. Census data and normalizes that data to maintain consistent Census tract boundaries across time. But while there is benefit to keeping census tract boundaries stable, Geolytics cost money and brought with it potential intellectual property restrictions and data discrepancies which could confuse future researchers who would use the more widely available original U.S. Census data.

The US Census data includes key variables that correlate to TOP's mission of bridging the digital divide. Census 2000 Summary File 3 (SF-3) tract-level data were downloaded, and thirty-one (31) measures concerning population, household, income, ethnicity and education were extracted for the entire United States. Age was eliminated from the process for the time being.

Two measures of income were extracted: per capita income and household median income. Poverty measures were not extracted, because the universe, or the denominator, in calculating all U.S. Census poverty measures is "those for whom poverty status is determined." The universe for other measures that were extracted is either total population or total households. In order to keep the demographic data layer simple, poverty measures per se were not used, but per capita income and household median income are measures which also indicate poverty in a given census tract.

SPSS and ArcMap's statistical functions were used to join the Census data with the ESRI census tract boundary shapefile.



Computer ownership statistics are of obvious importance for this project. The Current Population Survey (CPS), a monthly survey of 50,000 households which is carried out by the Census Bureau, regularly asks questions about household computer and internet presence and use. Data in the form of the CPS Table 1B, Presence of a Computer and the Internet for Households, by State was collected in September 2001, distributed in October

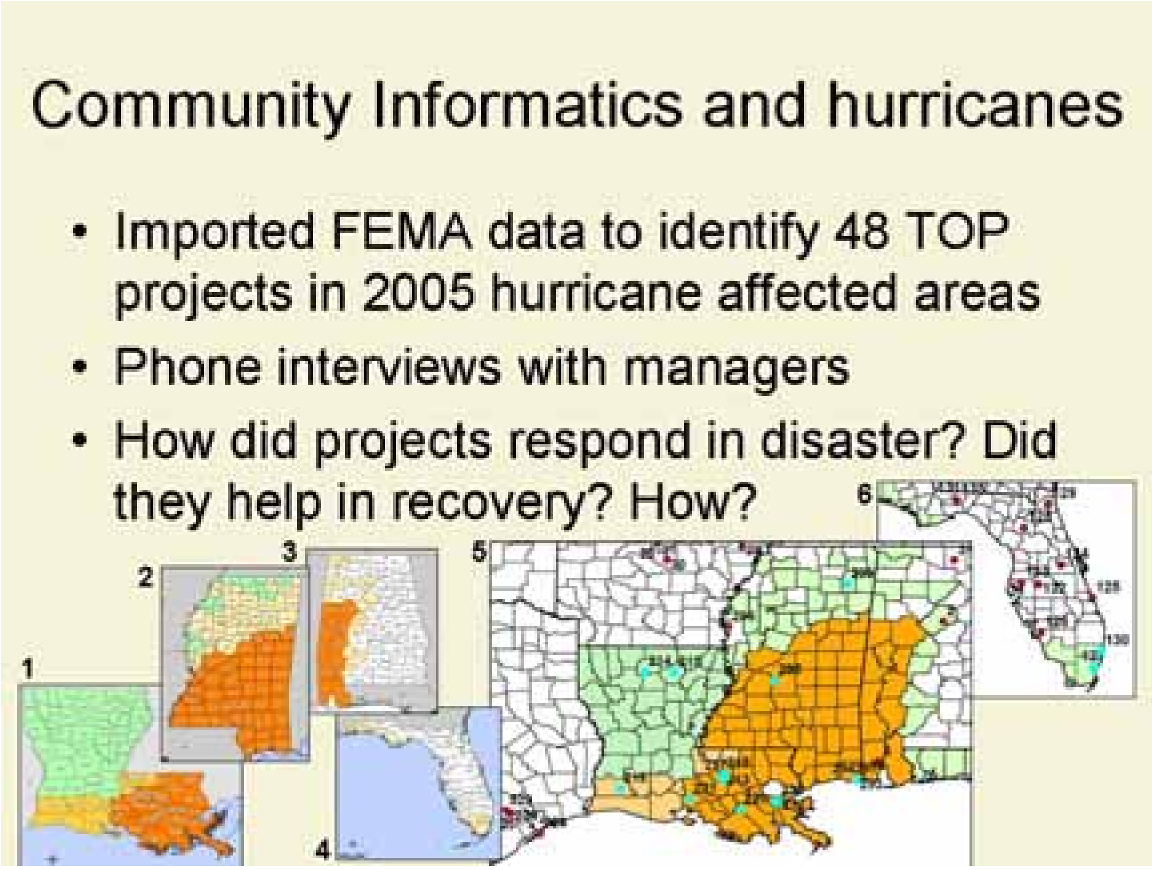
2003[1] and is seen here as a GIS layer, showing just one small example of the analytic capability of GIS data in combination with demographic data. The dots represent the TOP projects, and the shading of the states indicates how far the Internet has penetrated into households. The darker the state, the lower rate of home Internet access. In the states where fewer than 50% of residences have internet access, the dots show as red rather than black.

First, this map shows the breadth of the TOP program. It reached into every single state and it reached beyond the big cities in these states. And many of these dots represent projects that actually served multiple communities within geographic range of the project. This reveals the positive, democratic aspect of a public project. The benefit of archiving and studying such a public project is that we have data that sweeps across the entire country.

Second, it succeeded in reaching well into states—in a swath across the South—which are lagging in home Internet adoption. The southern states without large urban concentrations, without robust economic engines, were sites of multiple TOP projects, and these projects merit further examination. But in a general sense, we can see that TOP was indeed a very big bridge across the digital divide.

In total, the TOP data archive includes four data layers in its GIS maps: One layer contains data about the TOP projects. Another layer contains demographic data for the census tracts where TOP projects are located. A third layer contains demographic data for all the census tracts that are within one mile of a TOP project. And the fourth layer contains state-level data on computer ownership and use. It is important to note that this data does not necessarily correspond to the exact geographic service area of each TOP project. Future work can include mapping the service areas of TOP projects. There will also be the opportunity to build a layer of demographic data for buffer zones around each TOP project rather than relying just on the demographic data for the single census tract where a TOP project is located.

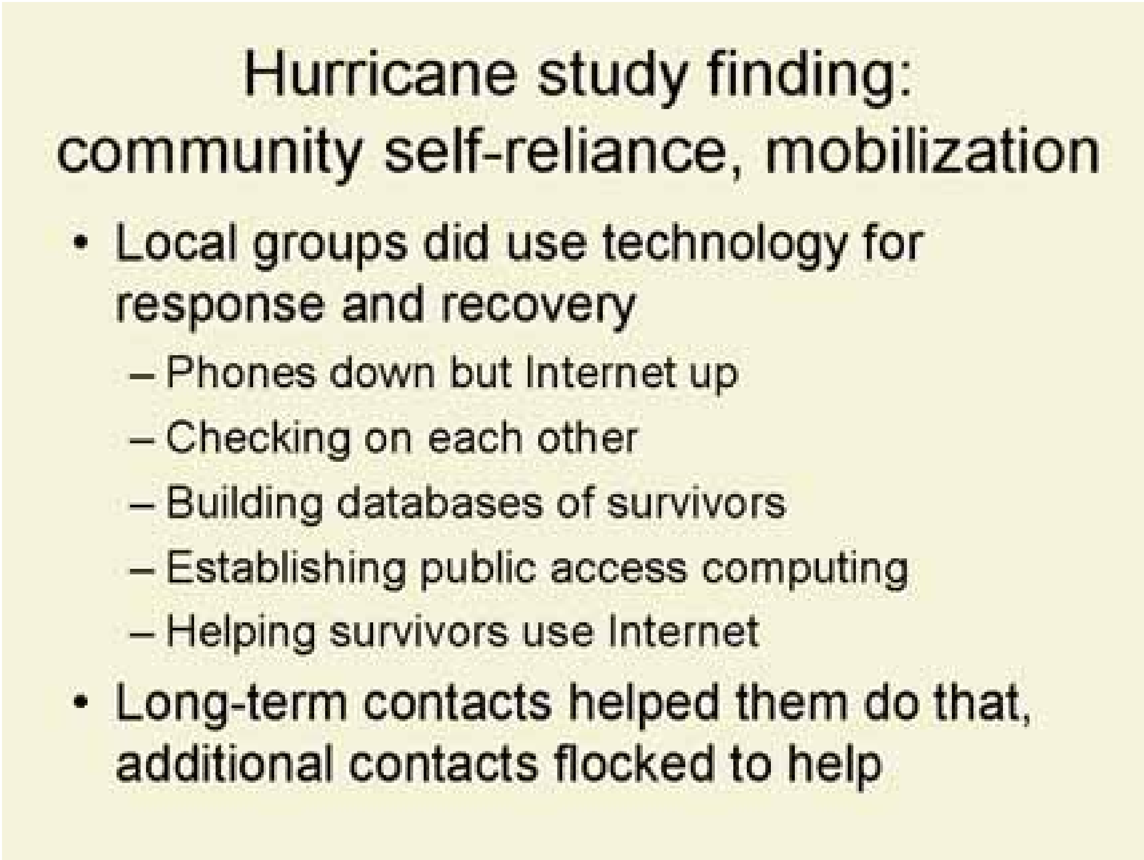
[1] The table was released on the Internet on October 27, 2005 and is available at http://www.census.gov/population/www/socdemo/computer.html.



Hurricane Katrina hit while the TOP archive work was underway. In reflecting on its terrible impact on local communities, we saw a need to assess the work of TOP in light of such crises. Hurricanes are just one type of disaster that communities face and it will help to know the role that local technology capacity plays in recovery. So we turned to FEMA, the agency that makes the county by county damage assessments that determine the type of recovery money that local people and organization can apply for.

Maps 1 through 4 in the slide above are examples of what is publicly available from the FEMA website. They show different levels of eligibility for assistance after Hurricane Katrina, so they give an indication of Katrina's destructive impact, with dark orange being the highest eligibility, light orange a medium level, and green the lowest level. Importing this FEMA data for every 2005 hurricane—manually—into the GIS files resulted in a set of maps including maps 5 and 6 above.

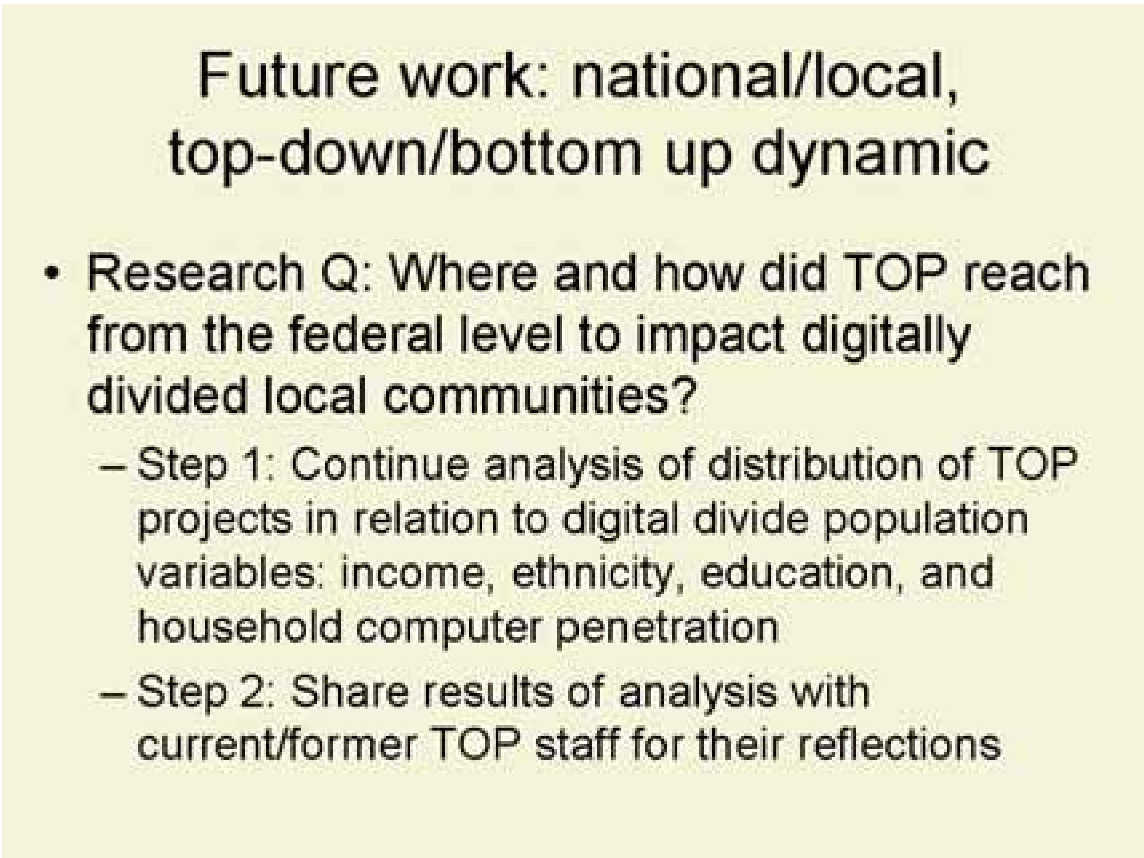
Across all the states and all the hurricanes, 48 TOP projects were found that could have been affected in some way by the hurricanes. They might have been directly in the hurricane path, or they might have been in areas that were strained or stressed to provide help and support, often long term, to adjacent communities which had evacuated. Team members called these projects and have carried out nine lengthy phone interviews with managers to obtain the details on how they were impacted, how they responded, how they played a role in community recovery.



Findings so far are somewhat surprising and quite hopeful. First, even under conditions where one might guess that information technology was inoperative due to electrical system failures, the failures were uneven and sporadic and where the technology could work, it was used. Electrical power tended to come back on while the communities as a whole were still quite devastated, as we have all seen on TV. In one case the Internet worked—via T-1 and cable networks—while the telephone network was still not functioning. So people turned to email to report in and check on others who were all quite local to them.

The Internet and such applications as email and databases were key tools in connecting people to each other—people were determined and often desperate to check on each other's safety. This involved individuals, families, and public agencies. They were also needing and using the Internet to connect with resources—be it FEMA itself or the bus maps for the cities they had landed in post-hurricane. So projects to establish computer access for evacuees gathered a lot of energy from TOP projects, whether it was in the Houston Astrodome right after Katrina or in neighborhoods where survivors settled in the months after.

Our particular question—can communities help themselves with technology, bootstrap themselves into the digital world?—led to the following answer. When it came to making technology work in hurricane-affected communities, the TOP projects relied on people they had known for a long time, in other words, on other local people. Plenty more people then flocked to help, especially in the larger crises and the larger projects, like the Astrodome computer assistance to evacuees, but local, long-time contacts, pulled things together in the first place. It's very encouraging to learn this—that communities are resourceful and self-reliant in such terrible moments, in making use of technologies that most people still think of as new and complicated and hard to make work.



We have plans for a great deal more work on our TOP Data Archive, by us and by other researchers. Where is this all going?

One question we want to investigate is this: how can a national effort generate local transformation? Community informatics has established that digital transformation really works when it is locally driven, answering real needs in local communities, yet in the case of TOP, and many similar countrywide efforts, we see national prioritizing and federally-determined objectives. So how and where was TOP able to effectively reach from the federal level to excite—or galvanize or get behind—movers and shakers in local communities?

In community informatics this is known as the problem of social change that is top-down versus bottom-up. It looks like TOP solved this problem and we want to understand how and why. So we need to examine these projects more closely and the census data will help us find out where TOP made inroads and where it did not. This will include looking beyond the TOP project addresses to what are often multiple or distributed service areas.

The fortuitous circumstance where the TOP staff has a positive interest in our findings means that we are going to see what GIS can help us uncover, then take that to a focus group (or two) of current and former staff, and get their reflections on what dynamics might be operating to create the patterns we uncover.

And, we might add, next summer is when the data, the entire archive, will be available to others, and we hope to be introducing it at the Third International conference on Communities and Technologies, held next year at Michigan State University (https://ebusiness.tc.msu.edu/cct2007). Please let us know if you are interested!

Bibliography

The following books and journals provide a broad introduction to the core literature of community informatics. Additional work which includes more detailed bibliographies is available at http://www.umich.edu/~katewill.

Day, Peter, and Doug Schuler, eds. *Community Practice in the Network Society: Local Action/Global Interaction*. London: Routledge, 2003.

Gurstein, Michael, ed. *Community Informatics: Enabling Communities with Information and Communications Technologies*. Hershey, Penn.: Idea Group Publishing, 2000.

Keeble, Leigh, and Brian D. Loader, eds. *Community Informatics: Shaping Computer-Mediated Social Relations*. London: Routledge, 2001.

Schön, Donald, Bish Sanyal and William J. Mitchell, eds. *High Technology and Low-Income Communities: Prospects for the Positive Use of Information Technology*, Cambridge: MIT Press, 1999.

Schuler, Douglas, and Peter Day, eds. *Shaping the Network Society: The New Role of Civil Society in Cyberspace*. Cambridge, Mass: MIT Press, 2004.

Smith, Marc A., and Peter Kollock, eds. *Communities in Cyberspace*. London: Routledge, 1999.

Information, Communication and Society

Journal of Community Informatics

New Media and Society