PUBLIC HEALTH GIS NEWS AND INFORMATION

June 1998 (No. 22)

Dedicated to CDC/ATSDR scientific excellence and advancement in disease control and prevention using GIS

Next Stop! **"GIS in Public Health"**
3rd National Conference

August 17-20, 1998

Online conference information and registration: see http://atsdr1.atsdr.cdc.gov:8080/GIS/conference/
Abstracts will be available for viewing by the end of June (link to 'Conference at a Glance')
****

Selected Contents: Conferences (p.1-2); News from GIS Users (pp.2-9); GIS outreach (pp.9-11); Reports (pp.11-13); GIS lecture (pp.13-14); SEER/privacy (pp.14-20); Web site of interest (p.20-21); Final thoughts (pp.21-22)

I. Public Health GIS (and related) Events

☞ **SPECIAL** presentation to CDC/ATSDR: "Mapping Housing and Related Data: A Demonstration of HUD's Community 2020 GIS Software," Richard Burk, Department of Housing and Urban Development, Wednesday **June 24**, NCHS auditorium, 2:00-3:00 PM, and open to all. Sponsored by the NCHS Cartography and GIS Guest Lecture Series program. **Please make envision arrangements now**. [Contact: Chuck Croner at voice 436-7904, ext. 146 or email cmc2@cdc.gov]

☛ **SPECIAL** presentation to CDC/ATSDR: "EnviroMapper GIS and Open Data Access: A Demonstration of EPA's Interactive Interface on the World Wide Web," David Wolfe, U.S. Environmental Protection Agency, Wednesday **July 15**, NCHS Auditorium, 2:00-3:00 PM, and open to all. **Please make envision arrangements now**.

☛ Annual Meeting of the North American Cartographic Information Society, October 7-10, 1998, Milwaukee, WI [Contact: Cynthia Brewer at voice (814) 865-5072 or email cbrewer@essc.psu.edu]

☞ First International Health Geographics Conference, October 16-18, 1998, Baltimore, MD [Visit IHGC website http://www.jhsph.edu/ihge or contact Ric Skinner at voice 610-965-7060 or wskinner@fast.net]

☛ 23rd Annual Meeting of the Association of Public Data Users (APDU) will be held October 25-28, 1998, Alexandria, VA [See http://www.psc.lsa.umich.edu/DA/APDU/1998/ or contact conference co-chair Ted Hull, National Archives and Records Administration at voice (301) 713-6645, ext. 253]

☞ First International Conference on GIS Education - GIS/Ed '98, October 29 - November 1, 1998 Ypsilanti, MI [See additional information at http://ceita.acad.emich.edu/gised and brief description in this edition]

☛ North American Meetings of the Regional Science Association International, November 11-14, Santa Fe, NM [See http://geog.arizona.edu/rsai98]

☞ Annual Conference of the American Water Resources Association, November 15-19, 1998, Point Clear, AL [Contact: S. Rocky Durrans at voice (205) 348-1710 or email rdurrans@coe.eng.ua.edu]

☛ Annual Conference of the American Public Health Association, November 15-19, Washington, D.C.

[Contact: Carroll Lewis at voice (202) 789-5620 or email carroll.lewis@apha.org, or visit http://www.apha. org/convention/index.htm

☞ 1998 National STD Prevention Conference, December 6-9, Dallas, TX [Contact: Jean Hogan, CDC/ NCHSTP, at voice (404) 639-8268 or email jxh9@ cdc.gov]

☛ 2nd Annual Crime Mapping Conference, National Institute of Justice (NIJ), December 10-12, Arlington, VA [Contact: NIJ's Professional Conference Series Contractor at (703) 684-5300 or email nijpcs@ilj.org]

☞ 1999 CDC/ATSDR Symposium on Statistical Methods, "Emerging Statistical Issues in Public Health for the 21st Century", January 27-29, 1999, Atlanta, GA [Contact: Brad Myers at bam6@cdc.gov or visit http://www.cdc.gov/ od/ads/sag]

II. News from GIS USERS
(Please communicate directly with colleagues on any issues)
A. General News (and Training Opportunities)
1. From **Russell Kirby**, University of Wisconsin Medical School: I don't know if I mentioned it, but my collaborators (Seth Foldy, MD, Medical Director and Interim Commissioner of the Milwaukee Health Department, Mike Barndt, Ph.D., Professor of Urban Studies at UW-Milwaukee) and I just received a $15,000 grant from the Perinatal Foundation to geocode and map linked birth-infant death certificates for the city of Milwaukee and study spatial patterns of racial disparity in low birth weight and infant mortality rates. We view this as an important first step toward operationalizing our concept of a multicenter, collaborative GIS-Health Milwaukee. [Contact Russ at voice (414) 937-5610 or email r-kirby@whin.net]

2. From **David Mark**, University of Buffalo (USGIS 1998 Congressional Breakfast): Dear UCGIS Delegates and friends- Just a quick note to let you know that the 1998 UCGIS Congressional breakfast appears to have been a resounding success. Thanks to hard work by Harlan Onsrud, Tom Palmerlee, and

many delegates and colleagues, we had a much improved turnout. The breakfast was attended by at least 80 people, including 2 Senators, 5 members of the House, and 36 congressional staffers, as well as about 20 people from UCGIS institutions and about 10 other guests. Our co-sponsors, Senator Susan Collins and Senator Pete Domenici, both attended in person and presented kind and well-informed remarks about GIS and research. I introduced the breakfast, and was followed by excellent presentations by Lyna Wiggins (Rutgers), Mike Goodchild (UC Santa Barbara), David Maguire (ESRI), and Harlan Onsrud (Maine). The question-and-answer period also seemed to go very well, and staffers from several key members were present. Congressman Jerry Lewis, whose District includes Redlands, California, was present, and he chairs the House Appropriations subcommittee that includes independent agencies such as NSF. With assistance from the research office of the University of Maine, concrete plans are under way to request increased funding for GI Science research for FY99, and if it is too late for the coming fiscal year, we seem very well positioned to work with both Congress and NSF on increased funding for the following fiscal year. Thanks again to all those who made this Congressional Breakfast what I think will be a major leap forward for our organization and for GI Science funding in the US. Be sure to follow up with your local members of Congress, and it is worthwhile letting your university or organization's administration know about it, too. More details should follow soon. [Contact: David at ucgis-president@ucgis.org; to check out the Congressional Breakfast agenda see, http://osu. orst.edu/dept/geosciences/congress/ congress 98.html]

3. From **Phillip Bouton**, National Association of County and City Health Officials (NACCHO): The National Association of County and City Health Official's Pollution Prevention Project seeks information on local health department programs or activities related to pollution prevention that use a geographic information system (GIS). This information will be used to develop case studies and demonstration of GIS applications at NACCHO's annual conference in September of 1998. The project

is also developing a GIS fact sheet that will answer basic questions related to start-up, cite available resources, and highlight potential uses of GIS applications using pollution prevention activities as an example. Pollution prevention activities are those that lead to: (a) reduction of the amount of any hazardous substance, pollutant, or contaminant entering any waste stream, or otherwise released into the environment prior to recycling, treatment, or disposal, and; (b) reduction of hazard to public health and the environment associated with the release of such substances, pollutants, or contaminants. Sites selected for case study will receive an expense paid trip to the NACCHO '98 Annual Conference to demonstrate their GIS application. Please send or call with information on GIS applications that specifically lead to pollution prevention activities and/or address pollution prevention. Include contact information, the type of application, purpose, outcome, and lead agency and other agencies involved. [Contact: Phillip Bouton at voice (202) 783-5550 or pbouton@naccho.org]

4. From **Cynthia Warrick**, Howard University: The Howard University Urban Environment Institute's Historical Black Colleges and Universities (HBCU) GIS Workshop will take place on July 26, 1998 through August 2, 1998. A group of about 7 teams (2 HBCU faculty + 1 leader/trainer) will spend 3 ½ days using GPS and perhaps digital photography to map important features (trails, major buildings, gardens) of Ft Dupont Park DC, 152 ha (376 acres) for an NPS Management Plan [Contact: Cynthia at voice (301) 593-5725 or email cwarrick@gmu.edu]

5. From **Jay Morgan**, Towson University: The First International Conference (GIS/Ed '98, see announcement Section I) follows a series of related international symposium on GIS in higher education held in Miami in 1991, in Columbia, Maryland in 1996, and in Chantilly, Virginia in 1997. The issues and action items of concern to all GIS educators raised in these symposia have led to the birth of GIS/Ed'98. The GIS/Ed '98 theme is "Expanding GIS Education Locally and Globally through Technology and Outreach." Once again, the purpose is to provide a forum for discussion of today's strategic issues in GIS Education. Abstract submission deadline is June 30, 1998 [Contact: Jay at voice (410) 830-2964 or email jmorgan@towson.edu]

6. From **Lois Dean**, HUD: A specialist meeting on Empowerment, Marginalization and Public Participation GIS (PPGIS) has been announced by Project Varenius, a new NSF-funded research initiative in geographic information science of the National Center for Geographic Information and Analysis at the University of California, Santa Barbara. The meeting will be held in Santa Barbara, California October 14-17, 1998, and proposals are due July 17, 1998. This Specialist Meeting is expected to bring together individuals who have extensive experience with PPGIS. It will be a forum for sharing experiences about alternative GIS designs and applications which better reflect community interests and involve and empower its members. The meeting will also be concerned with ways in which PPGIS can have unintended consequences by marginalizing people and communities. This initiative will, therefore, explore the contradictory nature of PPGIS design and implementation through presentations of case studies in a diversity of social contexts. A PPGIS research agenda will be developed and plans established for a possible subsequent conference. This followup conference would involve community groups, policy makers, planners, government agencies, NGOs, GIS vendors, private sector representatives, and academics, that are involved in PPGIS. [Contact: See Call for Proposals at http://www.ncgia.ucsb.edu/varenius/ppgis/call.html; also, If anyone would like to collaborate with Lois on a proposal using HUD's Community 2020 software, especially Version 2, please contact her at Lois_Dean@hud.gov]

7. From **Jack Eichenbaum**, NYC GIS Steering Committee (through ppgis-scope@igc.org): NYC City Council is holding hearings on GIS next week. One topic is the type of GIS organization city government should have. Should it be part of an existing city agency, a separate agency, a non-profit corporation or something else? Those in large cities or counties with

experience in this debate are sought for comments, advice. [Note: If you need to reply in confidence, reply to Jack at jaconet@aol.com]

B. Technical News

8. From **Tom Allnutt** (through ECOLOG-L): The USGS-Alaska Biological Science Center announces the availability of a software tool for the analysis of animal movements within an integrated GIS environment. The software, written by Philip N. Hooge of the Glacier Bay Field Station, is a collection of 32 functions that comprise an extension to the ArcView GIS program, which is available on Unix, PC or Macintosh platforms. Functions in the extension include: parametric and non-parametric home range models, a test for site fidelity, sample size tests, tests of complete spatial randomness, and functions to aid in path analysis, habitat selection and the GIS processing of locational information. The program and online documentation can be downloaded from the Internet at http://www.absc.usgs.gov/glba/gistools/index.htm. Also available is a package of raster GIS tools to extend the capabilities of the Spatial Analyst extension to ArcView. This program comes in two versions one requiring Spatial Analyst (full implementation) and the other requiring only ArcView (several functions disabled). The current version is 1.01 uploaded on 4/21/98. This version is still being tested so please use with caution and send all comments and bug reports to Philip_Hooge @usgs.gov.

9. From **Geoff Dean**, Resource Information Forester, Tasmania, Australia: Many of your readers will be interested to learn of the release of a new S-PLUS product. "S-PLUS for ArcView GIS" provides access to S-PLUS from within Arcview. It also, "... provides access to spatial statistics in S+SpatialStats, an add-on module to S-PLUS. With S+SpatialStats, users can easily access comprehensive spatial data analysis and spatial statistical modeling tools for geostatistical data, lattice data and spatial point patterns"; http://www.mathsoft.com/splus/splsprod/arcview.htm is the web address for technical information, with a link to the press release. [I have no commercial connection with Mathsoft, and make this referral on the grounds of technical merit; Contact: Geoff at voice +61 3 6434 3249 or email Geoff.H.Dean@north.com.au]

10. From **Kirk Nordyke**, GIS Coordinator, Wyoming Game and Fish Department (through FWIM-L @LISTSERV.VT.EDU): Kirk requested information on the topic "Hardware configurations for GIS." His questions and eight subsequent responses are instructive: "I am currently in the process of replacing an HP Unix workstation running Arc/Info with a PC or PC workstation with Windows NT running Arc/Info and ArcView. What do I want (or perhaps I'd better say, need) for hardware? I know this is somewhat of an ambiguous question, but really any advise would be appreciated by this hardware-illiterate. Additional information: My MIS people have a good idea for the basics; they're talking a CPU upgradeable to a dual pentium, 300MHz, mirror the drive holding the OS and application software, 128Mb RAM, 6Mb VRAM, plenty of disk space, SCSI devices, etc. We plan on using the HP monitor (A2094A) currently in use and buying a new graphics card. However, I just ran across a message from a user that knew of a compatibility problem between ArcView on NT and a Matrox graphics card that we had at one time considered but decided against (not because we knew about the incompatibility). In other words, if anyone knows of other compatibility problems, please let me know! Examples of working configurations would be great!

Also, I'd like to get an idea about what might be the best and/or most popular form of backup device nowadays, or is it not clearcut? I'm sick of tape devices - too slow and unreliable. What about rewritable CD-ROM, or optical drives, or the removable ZIP or Jaz drives?" Responses:

(1) We primarily use tape for storage only because of the capacity of the storage. Of course, tape cannot be used for quick access, it's more of just an archiving tool for retrieval of data when needed. We also have a CD writer (not re-writeable) that burns CD's at a relatively low cost ($10/CD), but they only hold 650meg per CD. Plus, they only write once, so they're not good for using as additional storage space. They're better for burning data to the CD and archiving it. We

have a few zip and jazz drives, but we don't use them all that often. This, however, is probably due to the enormous size of the data we deal with (several gigs at a time). Plus, I've found that these drives are also very slow. If you plan on storing a lot of data and accessing it often, the best bet is just to buy additional hard drive space. For archiving, we primarily stick with one-time CD's. And for personal data storage (backups and the like), we use 8 and 4mm tapes. A variety of storage devices increases the flexibility of the agency.

If you plan to have multiple PCs (now that you're getting rid of your UNIX workstation), you might consider having one, centralized drive that is shared across a network. We have both a UNIX and a PC lab that have access to drives located all over the network, but we're going to a centralized RAID system to handle all of our data sharing. If you have data stored on each PC, and people on other PCs need to get to it, you start running into a messy situation. A well-planned network can do wonders for efficiency.

(2) IF given a choice I would go with a Jaz over the Zip drive. The zip only has 100meg capacity so for many GIS coverages you can only back up one coverage. We have a Zip drive and can run into problems with space.

(3) . . . I set up the lab at the Coop Unit. What I ended up with was an ALR Dual 6 with a Pentium Pro 200 workstation (single processor, we didn't install a second), 64 MB of RAM, and a SCSI hard drive that was too small at 4 Gigabytes. We ran Windows NT 4.0 and the new version of Arc/Info, and it ran smooth as silk. We had a Matrox Millennium graphics card (2MB) that presented no problem for us, so I can't speak for any potential conflicts. I think that 300 MHZ and 128 MB RAM will serve you well. As far as backup goes . . . some read-write optical disks were (fairly) fast and held like 500 or 600 MB. I've been impressed with the JAZ drives; 1 gigabyte provides adequate backup space, and it reads and writes at a good clip (especially if you're used to waiting on tapes!)

(4) . . . we use a cheaper type Matrox Mystique Business 2000 card. The Milleneum cards are two to four times more expensive than this one and said to be one of the best cards. However I do not have any experience using them. If you read one of the text files released with ArcView you will find that ArcView has been optimized for S3 chip video cards. That may cause some problems with cards equipped with other chipset.

(5) My shop (Dept. of Natural Resources, Office of Resource Conservation, Watershed Management Section) has transferred our GIS from a SUN based UNIX box to a COMPAQ Professional Workstation about six months ago. I've been very happy with the results. Essentially, I'm running NT4.x variant of ArcInfo 7.1.1 and ArcView 3.0a on a 200MHz Pentium Pro processor (dual Pentium capable - we may add a second processor sometime this year) with 64Mb of RAM, 8MB of video ram, 4Gb hard drive, CD-ROM, 1Gb external IomegaJaz drive (2Gb Jaz drives are now available) and a 21" monitor. This configuration is adequate for our needs at this time. As we get more involved with DRGs, DOQs and DEMs, I'd like to double the RAM and install the second processor. The Jaz drive has worked out very nicely as a backup and external storage device.

(6) I'm running ArcView on a HP P200 Vectra XW, 256 MB RAM, 2 - 2GB Ultra SCSI HD's and an Accel 3D video card that came standard with the PC. This PC works very well with ArcView, with the ability to load over 1 GB of themes into a project and still display everything within three minutes. Producing 300 MB of themes with labels and complex symbology takes less than a minute to display. We buy standard configurations because we buy so many, and we can lay the responsibility for failures or incompatibilities at the manufacturers' feet. We usually up the RAM before upgrading anything else, and while 128 MB will work, if you came from UNIX you'll want 256 MB or more to get performance like you're used to. In my next PC, I'm getting 512 MB RAM, mostly for ArcView speed (by not paging to the HD).

Rewritable CD-ROM is fine, but requires use of the PC for nothing else during CD creation. Otherwise, PC's can burp and ruin the CD. I use a Pinnacle Micro Vertex 2.6 GB magneto-optical drive that formats the media to 1.06 GB each side, using NTFS with 512 byte sectors. MO media have a stated 30-year life,

which approaches CD's. The media have been reliable for me, although I would never buy a Pinnacle Micro product again; the drive failed twice and took several weeks to be repaired each time. There are several other manufacturers including Sun and Sony that make a better drive. I use both JAZ and ZIP drives, and while they are convenient I have had files corrupted in one day. These are not good archival media, but can work fine for daily and weekly backups.

(7) I don't know about the Jaz drive, but I just got a HP SureStore DAT24 tape drive, and it's great. It holds 12Gb uncompressed, and 24Gb with 2:1 compression. It's really fast, too! I can back up about a gigabyte in 12 minutes. It has a SCSI interface, and it's so quiet you don't even hear it running. Kind of pricey, though; I think we paid about $1200. I also have a parallel port zip drive that has worked fine with light usage, but it's really slow and there have been murmurs of a reliability problem (the click of death). Next thing is to get a Jaz drive. If you get one, can you post how long it takes to backup a gig?

(8) We have an external Zip drive that's been used mostly for slides scanned at another location and brought here to use with ArcView. Most of the newer machines have internal tape backup units (I think we have an external too), but the GIS people agree with you that they are too slow. I have not experienced tapes as unreliable, at least with the ones I used in my former position. Though the tape drive on the less-than-one-year-old server here hasn't worked since last fall, and I haven't been able to get it to work either yet. We also have a couple of external SyQuest drives that were purchased basically to be used as additional "hard drives" for laptops, before Zips were big enough. One of the current users of the SyQuest is not real pleased with it, but I have no experience with it.

We just got, as you may have noticed on the listserv, a HP SureStore 7110e CD writer (for CD-R and CD-RW) primarily for the GIS people to make temporary and permanent backups and to free up some hard disk space. We haven't had much luck with that yet. Yesterday we were on the phone with HP and couldn't reboot to Win95 after making suggested changes and ended up reformatting the C: partition and reinstalling Win95 (HP said we had a virus, we had no

troubles until we made their suggested changes). From the responses I've gotten from my listserv posting, it sounds like the technology is not very consistent yet. A common problem is that the CD's are not universally readable. Note that RW CD's only hold about 500 MB (at least with this drive) on a 640 disc due to formatting. Also, I don't think a RW disc can be read in a CD drive that is not RW.

11. From **Chris Austin**, GeoHealth Inc.: I would like to make you aware of our new Health care/GIS product, BodyViewer for ICD-9 Codes. In a nutshell Body Viewer for ICD-9 Codes uses digitized images of human organ systems to link clinical health care data (ICD-9 codes) to the power of a GIS. This lets the user analyze and map patients by disease type and even symbolize the patients by some attribute. For instance, a user could map all patients with digestive disorders and then each patient could have a different colored symbol according to their type of employment. Also, watch for our latest product PatientAccess. This product will be available in July/August. This product is designed to answer questions like, "how many patients are within "X" miles of a health care provider network?" These two products are designed to work separately or work together. An example of how they might work together is a health care entity uses BodyViewer for ICD-9 Codes to analyze a patient population and map all the patients that have cardiovascular disease. This cardiovascular patient population could then be analyzed using PatientAccess to see how many of them are within 20 miles of a Cardiologist or cardiovascular facility. [Contact: Chris can be reached at caustin@geohealth.com or visit GeoHealth Inc.'s web site at www.geohealth.com]

12. From **Bernard Guenette**, Synoptech Inc.: We have just released our cartographic software application 3D Geographer™. We believe this product is of interest to your readers. 3D Geographer™ builds, displays, animates, and edits 3D geographical scenes efficiently and transparently. It is a powerful yet easy to use map builder that runs on Windows 95/NT. 3D Geographer™ imports and combines data from 11 geospatial formats--up to 2,200 vectorial layers of

information. Using this data, 3D Geographer™ can create more than 25,000 complex 3D geographical scenes. 3D Geographer™ also integrates 2D and 3D CAD models. 3D Geographer™ supports these data formats: USGS:DEM, USGS:DLG, USGS:LULC, USGS:SDTS, NTDB:CCOGIF (Canada), DCW, DXF, and DGN. 3D Geographer™ also includes ETOPO5 and the CIAWDBII as customized 3D atlases. You can visit us at: http://www.3dgeographer.com. The price for a single licence of 3D Geographer™ is $699.00 U.S. (plus shipping and handling). A 50% discount is applied to orders from non-profit organizations. [Contact: Bernard at voice (514) 847-0047 or email bguenette@ synoptech.ca]

C. Internet News

13. From **Lois Dean**, HUD (through ppgis-scope @igc.org): A group of Federal agencies and representatives of state and local government are working on standards to make it easier to transfer geographic data electronically. They welcome comments and volunteers to review proposed standards. A new FAQ page is now online for the US Federal Geographic Data Committee's (FGDC's) Spatial Data Transfer Standard (SDTS). Please note that the new Frequently Asked Questions (FAQ) page is now on line at: http://mcmcweb.er.usgs.gov/sdts. This new FAQ covers a great deal of territory and includes all the updated information available at the time of release. It also covers a number of the more common questions asked of the SDTS mailing list. [Contact:Dan Henke, SDTS Task Force, at sdts@ usgs.gov]

14. FGDC Standards: The FGDC develops geospatial data standards for implementing the NSDI, in consultation and cooperation with State, local, and tribal governments, the private sector and academic community, and, to the extent feasible, the international community. Anyone interested in participating in any of the FGDC standards activities is invited to contact the Chair of the sponsoring FGDC Subcommittee or Working Group. The FGDC Standards Reference Model, defines the expectations of FGDC standards, describes different types of geospatial standards, and documents the FGDC standards process. The Standards Directives provide additional guidance to the FGDC Subcommittees and Working Groups developing standards and document the practices of the FGDC Standards Working Group. Note many of the documents on this page are in Portable Document Format. You can download a free PDF reader from Adobe. For more standards information, go to the FGDC Standards information site at http://www.fgdc. gov.

The Subcommittee on Cultural and Demographic Data (SCDD) of the Federal Geographic Data Committee (FGDC) is composed of members from Federal agencies that collect or finance the collection of cultural and demographic geospatially referenced data (public health included) as part of their mission or that have direct application of these data through legislated mandate. The SCDD's purpose is to promote the collection, use, sharing, and dissemination of cultural and demographic geospatially referenced data pertinent to the needs of the Nation. These data include the characteristics of the people, the nature of the dwellings in which they live, the economic activities they pursue, the facilities they use to support their health and recreational needs, the environmental consequences of their presence, and the boundaries, names and numeric codes of geographic entities used to report the information collected.

The SCDD engages in the following activities: Developing and maintaining governmentwide standards for the description, documentation, and exchange of geospatially referenced cultural and demographic data; and, identifying key cultural and demographic geospatial data sets that meet critical national planning and management needs and publishing classification schemes for these data. The subcommittee works to improve access to these data within the restrictions imposed by applicable confidentiality and privacy regulations. [Reference: see site http://www.census.gov/ftp/pub/geo/www/ standards/scdd/]

15. GIS and Privacy (through **Pat Vaive**, NCHS, and **John Fanning**, OS, ASPE, DHHS): Information and Privacy Report by Commissioner Tom Wright,

Ontario, Canada on the topic Geographic Information Systems, April 1997. From the Introduction- "As government organizations move with increasing speed toward the creation of a predominantly electronic environment, it is not always clear how to facilitate these purposes and rights. New questions and concerns inevitably arise about the protection of privacy, the nature of public access to information under the control of government organizations (government information), and the applicability of Ontario's freedom of information and protection of privacy legislation. The access and privacy issues associated with today's electronic environment, of which GIS technology is a part, are complex and challenging. [For the full report, see: http://www.ipc.on.ca/web_site.eng/matters/sum_pap/papers/gis.htm#; Upon request, this publication will be made available on audio tape]

16. National Library for the Environment: The on-line National Library for the Environment http://www.cnie.org) contains seven free and very useful information resources. The Library is heavily used by students and educators, NGOs, journalists, and decision makers in the private and public sector all over the U.S. and around the world. The Library was recently honored with a Computerworld Smithsonian award for innovative use of technology. These resources include:

a. Over 300 nonpartisan environmental reports produced by the Library of Congress' Congressional Research Service exclusively for Members of Congress and their staff. Issues addressed include: Agriculture & Grazing, Air Issues, Biodiversity, Climate, Energy, Forestry, International Issues, Legislation, Marine Issues, Mining, Natural Resources, Pesticides, Pollution, Population, Public Lands, Regulatory Reform, Stratospheric Ozone, Trade, Taxes & Economics, Transportation, Waste Management, Water Quality and Wetlands.

Reports are reviewed for technical soundness, objectivity and nonpartisanship and many are updated monthly. Yet, surprisingly, these reports are not available to American citizens. As the National Journal reported, "[i]f anyone other than a Member of Congress asked the Congressional Research Service

for copies of the reports the agency writes, the answer would be an emphatic "no." The Library of Congress is not affiliated with the National Library for the Environment and does not cooperate in this initiative. However, as products of Congress, CRS reports are not copyrighted and thus CNIE has legally made them freely available to the public. [See http://www.cnie.org/nle/crs_main.html]

b. Environmental Education Programs and Resources include: Directory of Higher Education Environmental Programs; Starfish: sustainability courses, bibliographic references, innovative teaching techniques; Academic Programs in Conservation Biology; Community college and high school environmental technology programs; and, Environmental Impact Assessment Training Courses.

Environmental Education Programs and Resources are implemented in collaboration with Rice University and include resources developed by us and others. The Directory of Higher Education Programs includes over 150 degree-granting, environmental programs -- many of which train K-12 teachers. The Directory was developed by CNIE and Rice University. The resources provided on Advanced Technology Environmental Education Center (ATEEC) in Iowa address environmental technology education through curriculum development, professional development and the nation's community college and high school environmental technology programs. Complementing the program-focused resources is Second Nature's Starfish, with over 200 syllabi and reading lists for sustainability courses, 1600 bibliographic references for sustainability and 21 innovative teaching techniques. [See http:// www.cnie.org/educate.htm]

c. Population and Environment Linkages contain extensive information and source documents at all levels of complexity in an innovative and simple framework. Areas addressed include: Demographics, Fresh Water Resources, Oceans, Land-Use, Coastal Environments, Air, Climate, & Atmospheric Change, Food Resources, Biodiversity, Security, Development and Economics, and Environmental Health. Abstracts and full bibliographic information for each article is given. Introductory articles on how human population impacts upon aspects of the environment can be found

on the population/environment database home page. [See http://www.cnie.org/pop/pophome.htm]

d. Environmental Laws and Treaties is maintained by Pace University Law School and includes organized links to primary legal sources on: International Laws and Treaties, Federal Environmental Laws of the United States, State Environmental Laws, and Comparative Environmental Legislation [See http://www.cnie.org/ nle/nlelaw.htm]

e. The Virtual Library of Biodiversity, Ecology and Environment is maintained by Rice University Center for Conservation Biology Network and organizes information around the following topics: Global Sustainability, History of Life, Endangered Species, Captive Breeding, Exotic Introductions, Pollution, Protected Areas, Values of Biodiversity, National issues (non-U.S.), State Issues (U.S.), U.S. Government and Legislation, International Treaties, Biodiversity and Conservation, and Conservation Education [See http:// www.cnie.org/biodi/bioframe .htm]

f. Meetings and Conferences [See http://www.cnie. org/conferences.htm]

g. Careers and Opportunities includes: Environmental Positions, Counseling, Corporate Research, Job Market Analysis, Salary Determination, Job Databases, E-Mail Headhunters, Resume Preparation, Resume Posting, Cover letters, Interviewing, Education, and links to other career sites (Mega Lists). [See http://www.cnie.org/career/mega job.htm]

All of these sites are accessible from the home page of the CNIE: http://www.cnie.org [Contact: Kevin Hutton, Webmaster, Committee for the National Institute for the Environment at voice (202) 530-5810 or email khutton@ cnie.org]

17. Forwarded by **Donna Higgins**, EPO and **Bobby Milstein**, OPPE (from the Loka Institute): The Silicon Valley Toxics Coalition -- which has fought high-tech pollution for over 15 years -- is using the high-tech industries' own tools and technology to map the industrial pollution in Santa Clara County. In celebration of Earth Week, the Silicon Valley Toxics Coalition (SVTC) has added a new feature to its website at http://www.svtc.org/resource.htm. This site contains maps and data from EPA and the Regional Water Control Board about the 178 groundwater contamination sites in Santa Clara County, including Superfund sites. It also enables people to compare the worst pollution sites with demographic information, including census data on race and economic status and children.

We worked with Mike Meuser, who helped develop the Environmental Defense Fund's new scorecard website maps, and local geographic information system consultants Michael and Francie Stanley-Jones. SVTC is now developing maps which in a point and click format will provide information about the 178 groundwater sites, the extent of contamination, and information about groundwater plumes.

This project goes beyond the maps done by the Environmental Defense Fund, by enabling people to view pollution and census data at the neighborhood and street level for Santa Clara County. These maps highlight the connection between environmental pollution and environmental injustice." Through this GIS mapping project, SVTC plans to integrate the information provided by various agencies and neighborhood demographic data with maps of Santa Clara County. With this information communities will be able to use the maps to link documents and information to make sense of environmental, social and economic decisions, and to increase community participation in those decisions that affect the way people live their lives.

SVTC's mapping project is a work in progress. SVTC is planning more maps which will include information about Toxic Release Inventory (TRI) releases--not just air and water but also transfers and storage from the facilities that reported under EPA's Toxic Release Inventory in Santa Clara County. [Contact: LOKA Institute at crescent@reeusda.gov]

III. GIS Outreach
(Editor: All solutions are welcome and will appear in the next edition; please note that the use of trade names and commercial sources that may appear in

Public Health GIS News and Information is for identification only and does not imply endorsement by CDC or ATSDR)

☎ From **Sam Posner**, CDC/NCCDPHP: We have HIV prevalence data from 1988 to present from blood banks in northern Thailand. We are interested in using GIS to describe the geography of the epidemic. Rather than use all the data we were going to use data from January of every year. The data includes complete address including post code. Do you know if the international maps that are available for use with GIS software can match post code in northern Thailand? What would be a good intro book on this? Is post code the appropriate unit for geocoding? If not what would you suggest using? [Contact: Sam at voice (770) 488 5060 or email shp5@cdc.gov]

☎ From **Mark Hill**, University of Pennsylvania: I have just purchased the most recent version of the Social Security Death Master File for a research project. I wonder how important the Social Security Death Master File is for epidemiological research. If you have used these data in your research or have intentions to do so in the future, please send me a note (describing how it was used). If you don't know what the SSA DMF is, check out a summary at http://www.ancestry.com/ ssdi/article.htm. [Contact: Mark at voice (215) 898-3166 or email mhill@ pop.upenn.edu]

☎ From **Mimi Roddy**: I am a student working with Dr. Jim Vanderslice on GIS projects at the University of Texas- Houston SPH in El Paso. We are interested in finding out more about the extent and types of GIS applications that are currently being carried out by public health practitioners in county health departments and state departments of health. What kinds of projects are practitioners pursuing and what software is commonly used? Where do they find spatial and health outcome data? Any help or guidance you can give us in our pursuit of this information is greatly appreciated. [Contact: Mimi at email mroddy@zianet.com]

☎ From **Lissa K. Blash**, San Francisco State University (GIS and Welfare to Work): I would be interested in hearing about any work people are doing using GIS for Welfare-to-Work planning. Is anyone out there using GIS to look at service provision, transportation, job growth sites, etc.? [Contact: Lissa at lblash@sfsu.edu]

☎ **Joe LeMaster**, University of Minnesota: Can anyone give specifics about contacting Epi Info by email (i.e., an email address from which they have actually received an answer) and also information about the next version of Epi Info. I was under the impression that we would not see any more versions. I was weaned on Epi Info as an MPH student, but eventually began to use STATA. In recent years, I have been unable to get any responses from the Epi Info team, but I have not had email. Is there an Epi Info email discussion list, like there is for STATA?

Response: Have you tried CDC's Epi Info web page, http://www.cdc.gov/epo/epi/epiinfo.htm? Also, CDC provides technical support for the Epi Info, Epi Map and SSS 1 programs. This support is free but is limited to problems that cannot be solved after reference to the respective manuals. Hotline support is available in English and Spanish. The hours of operation of the telephone support are Monday-Friday, 8am-5pm (EST). To contact the Epi Info hotline:
1. Telephone: (404) 639-0840 *Support is limited to 30 minutes per call.*
2. Fax: (404) 639-0841
3. Email: epiinfo@cdc.gov

There is also an Epi Info discussion group. You can subscribe to it on the WWW at http://www.cdc. gov/ epo/epi/ epimail.htm. This is what they had to say about the new version: What's Going on With Epi Info 2000, the Windows Version of Epi Info? The Epi Info programming staff is hard at work on the next version of Epi Info. A Demo version is completed. It is designed for Windows95 and WindowsNT, with the following main goals:
1. Maximum compatibility with industry standards, including: Microsoft ACCESS and other SQL and ODBC databases

 -The Visual Basic for Applications programming

language (with extensions to allow as many of the user-friendly Epi Info commands as possible)

-World Wide Web browsers and HTML

2. Extensibility, so that centers outside CDC can produce additional statistical modules using the core programs of Epi Info 2000 for data entry and management

3. Ease of use

4. As always, distribution in the public domain

For efficiency and compatibility, and to be able to keep up with the frantic pace of development of programming tools (particularly the World Wide Web), we are using Visual Basic as the development language, with liberal use of commercial programming "controls" for functions such as mapping and graphing that require extra speed. Database management is via the SQL language and the Microsoft JET Engine.

Epi Info 2000 is not just a "port" of Epi Info 6 to windows but an entirely new set of programs designed around these concepts. Epi Info 2000 will provide Epi Info's strengths (e.g. instant creation of a database from a questionnaire and simple commands like LIST, TABLES, and FREQ) along with the graphics and printed output of Windows, and compatibility with larger databases likely to be found in the hospital and public health databases of the future. The effort in Epi Info, Versions 5 and 6, to create an open system for contributions by other programmers, will be extended to allow for varied statistical routines and other specialized functions to be produced internationally.

In early December 1996 the core programs took shape. These are MAKEVIEW (Questionnaires are called "Views" in Epi 2000), ENTER, and ANALYSIS, as well as an early versions of the graphing module and main menu. Development of a manual, other functions like mapping and graphing, and the programmability for which Epi Info 6 is known will take a few more months, but a Demo version of Epi Info 2000 is available via FTP. I hope this helps. Good luck! [Source: S. Simons, Jazminetree@mindspring.com]

IV. Special Reports
(Submissions are open to all)

[Editor: The following **two funding opportunities pertaining to GIS in cancer epidemiology** were received from the National Cancer Institute, Epidemiology & Genetics Program, Division of Cancer Control and Population Sciences. For complete information (excerpts only are shown below) on RFA CA-98-017 or RFP/NO2-PC-85074-39 please contact Theresa Shroff, Contracting Officer, at voice (301) 435-3796 or email ts144t@nih.gov]

A. TITLE: Regional Variation in Breast Cancer Rates in the U.S. RFA CA-98-017; Letter of Intent Receipt Date: June 30, 1998; Application Receipt Date: August 25, 1998. RESEARCH OBJECTIVES: Geographical variations in cancer rates have been observed for decades, and described spatial patterns and trends have provided clues for generating hypotheses about the etiology of cancer. In the case of breast cancer, investigators have demonstrated that some variation can be explained by differences in the population distribution of known breast cancer risk factors such as menstrual and reproductive variables. However, regional patterns may also reflect an aggregate of diverse factors including, for example, varying presence of hazards in the environment, demographics and lifestyle of a mobile population, subgroups of susceptible individuals, and changes and advances in medical practice and health care management. Disentanglement of these factors is necessary to assess associations, singularly or jointly, with breast cancer risk in individuals and populations.

It is unclear what the determinants of geographic differences of breast cancer rates may be and whether spatial variation of environmental factors is contributory in areas with elevated breast cancer incidence and mortality. Biologic data relating environmental pollutants to breast cancer risk are sparse, and epidemiologic studies have been challenged by methodological limitations, most often in determining past exposure levels. To date, the scientific literature on the association of measurable exposures, e.g., organochlorine pesticides, and breast cancer is conflicting.

Based on the known importance of endogenous hormones in breast cancer development, a leading question remains whether environmental factors, such as xenoestrogens and other hormone-mimicking pollutants, may also exert an effect. A multidisciplinary workshop, "Hormones, Hormone Metabolism, Environment and Breast Cancer," convened by the National Action Plan on Breast Cancer (NAPBC), the NCI, Tulane University, and the U.S. Public Health Service's Office of Women's Health, in September 1995 discussed the complexity of factors, unresolved controversial issues, and the need for improved methodology to measure hormones and their metabolites. The power of molecular and bioinformatic technology could potentially provide biologic probes and sensitive methods for epidemiologic studies to gain insights into the relationship between environment, the individual, and breast cancer.

RESEARCH SCOPE AND GOALS: The purpose of this RFA is to stimulate innovative epidemiologic studies that include assessment of markers or indicators of exposures, susceptibility or other factors relevant to human breast carcinogenesis. Major consideration will be directed to studies, including those that are transitional (from laboratory-based to population-based), that incorporate validation of utilizable markers, e.g., hormone-related, in human populations. Collaborations among multiple disciplines and research institutions are particularly encouraged, and research designs can make use of existing resources, such as specimen repositories. Supplementary research to expand an ongoing investigation (i.e., parent study) may be proposed, contingent upon the continuation of the parent study for at least two years. There is special interest in understudied populations, particularly those subgroups with unusually high breast cancer incidence and mortality rates, and in study populations of contrasting risks.

Investigators are encouraged to involve appropriate community/advocacy groups interested in breast cancer research. These groups could be comprised, for example, of breast cancer survivors, health care professionals involved in breast cancer care or women at high risk of the disease. The type and degree of participation by the group members could vary depending upon the proposed research activities, e.g., members could serve as advisors to the investigative team or assist in research implementation such as informing and recruiting eligible study participants in the community.

Studies responding to this initiative may propose, for example, research in areas listed in (but not limited to) this announcement (not shown in its entirety here), which includes: Application of computer technology, e.g., **geographic information systems**, and development of innovative statistical methods for improving estimates of historical environmental exposures.

B. TITLE: Geographic Information System (GIS) For The Long Island Breast Cancer Study. Solicitation Number: RFP/NO2-PC-85074-39; Issue date: mid-April, 1998; Proposals due: August 28, 1998. SYNOPSIS: The Division of Cancer Control and Population Sciences (DCCPS), National Cancer Institute (NCI) [has released] a request for proposals (RFP) for the Phase I Implementation of a geographic information system (GIS) to support the Long Island Breast Cancer Study Project (LIBCSP). The LIBCSP is a multi-study research effort in response to Public Law 103-43 (June 10, 1993) and consists of more than 10 epidemiologic and other research projects designed to investigate the relationship between environmental factors and breast cancer on Long Island, New York. The objective of this current project is to develop, test and document a prototype health-related geographic information system (GIS-H). The GIS-H will be used in support of investigations of contamination in Nassau and Suffolk counties in New York. The system will provide statistical and spatial analysis tools to support both the research and presentation/publication of results by graphical and cartographical display and to provide an efficient means for researchers to study disease incidence location in relation to location of potential hazard sites or environmental contaminants of interest, both current and past. Services and materials to be provided shall include, but not be

limited to 1) GIS software; 2) GIS data conversion services; 3) GIS database development to support GIS-H, including decision support for the NCI-specified functionality; 4) Computer hardware to operate the GIS software; 5) Site preparation, delivery, installation and testing; 6) Training; 7)Documentation and 8) GIS-H operation, maintenance and support. Phase I will consist of 24 months during which the system will be developed, tested and delivered. Phase II will consist of 3 option years for operation, maintenance and support. It is anticipated that the GIS-H implementation will consist of a cost-reimbursement, completion contract with options for 3 additional years. The SIC code is 7373.

V. NCHS Cartography and GIS Guest Lecture Series

(This section may include literature citations, abstracts, syntheses, etc., and submissions are open to all)

Mapping Housing and Related Data: A Demonstration of HUD's Community 2020 GIS Software," Richard Burk, Department of Housing and Urban Development, June 24, 2:00PM, NCHS Auditorium.

Abstract:

Public Health officials and researchers have traditionally analyzed the associations between location, environment and disease. We have come to accept that a variety of public health outcomes are associated with housing and population characteristics, e.g., lead exposure (age of housing), crime (low income housing), summer heat fatalities (poverty), radon (construction), and so on. These linkages come alive when the data are displayed on maps. In this day of rapid technological innovation, we are able to capture, store, process and display an increasing, unprecedented amount of information about ourselves and the places we inhabit. The challenge we face is how best to use that information; how to improve the way we turn the raw data into understandable information, make sense of it, and communicate it to others for better decision-making.

The applications that will be possible with broad, easy to use access to geospatial information are limited only by our imagination. We can get a sense of the possibilities by looking at some of today's applications of GIS and data that have spatial components. We have an unparalleled opportunity to turn the flood of raw data into understandable information about our society and the factors that impact on our health and well-being. This data will include not only digital maps, and economic, social, and demographic information, but also new data that we generate as we discover new relationships among these factors. If we are successful, the innovations in information technology will have broad societal benefits in many areas of decision-making for a sustainable future. One area of innovation is in automated systems for the capture, storage, retrieval, analysis and display of spatial data. The US Department of Housing and Urban Development through a joint venture with Caliper Corporation, developed Community 2020 as a full-capacity GIS tool for empowering people to analyze data and communicate its message on vivid, easy to manipulate maps.

Community 2020 Version 2.0 is a new easy-to-use software tool. For our clients, we loaded a high-end GIS software product, Maptitude, with HUD's project data for every jurisdiction in the United States and territories where we have program activity of any kind. The software is by no means limited in its application to HUD's concerns for housing and urban development. Users will find that they can input health data sets, ask and obtain instant answers to questions about the association of public health concerns with a multitude of covariables.

HUD enriched the already extensive database in the software, developed several user interfaces with Caliper Corporation to demystify the use of GIS, and has made it accessible to persons without special training or skills in the technology of mapmaking. HUD loaded the software with extensive Tiger Files, 1990 Census files (over 640 Census data categories) with 180 categories projected to 1997 and estimated for 2002 and 2007; Census data are available for Congressional Districts, ZIP Codes, Block Groups, Census tracts, states, counties, and many other jurisdictional boundaries. In addition, Census and other data are automatically calculated for any polygon or shaped area drawn by the user. Comparisons between cities or other jurisdictions are easy to make, as the criteria for a map in one location may be used

for any other simply by entering another address, ZIP code or other geographic information such as county or city. [Richard Burk is Director, Division of Community Viability, at HUD. He will demonstrate HUD Community 2020's many special features of interest to CDC/ATSDR and describe some of the plans for the future use of this and other technology to support HUD's mission. Contacts: Richard at voice (202) 708-2504, ext 4440 or email Richard_Burk @hud.gov; Lois Dean, Division of Community Viability, HUD, may be reached at (202) 708-2504, ext 4443 or email Lois_Dean@hud.gov]

VI. Related Census, DHHS and Other Federal Developments

Subcommittee on Privacy and Confidentiality of the National Committee on Vital and Health Statistics (Roundtable Discussion): Health and Medical Registries, Excerpts, Thursday, January 29, 1998, Hubert H. Humphrey Building, Robert Gellman, J.D., Chair (Note: This transcript is unedited - for the full transcript, see http://aspe.os.dhhs.gov/ncvhs/980129tr. htm). Mr. Gellman: Today's topic is registries. How do registries fit into the general scheme? My perspective is always that of the legislative scheme, but we are not specifically looking at any of the bills today.

What is a registry? A registry is a list of people. A list of students at school is a registry if you like. A list of patients in a clinic is a registry. A list of folks with a drivers license, which has health codes on it by the way, is a registry. A direct marketer's list of people with arthritis is a registry. It depends on how you define registries. How do we distinguish between a good registry and a bad registry, if those are appropriate adjectives? Who is entitled to open a registry and collect health data? What are the terms under which they collect the data, use the data, store the data? Those are some of the subjects today.

Dr. Brenda Edwards, NCI: First of all, the National Cancer Institute has developed a surveillance program. At the center of that program is SEER. It stands for Surveillance, Epidemiology and End Result. I think if you look at each of those components in our title, it actually characterizes both our roots and our future in terms of the kinds of activities in which we're engaged.

First of all, I can say that cancer registries build upon state-based and regional registries. The data we have in our federal database contains no personal identifiers. It contains a study or case I.D. So that any information I'm talking about in the SEER program, and that includes information that we make available through public use files on CD ROMs, including the data which has front end analytic packages; and you'll see on our Web site that you can make requests for our public use files; we're now trying to make that data more available, usable, and available for analysis and use by many people in many sectors.

So our data does not contain any personal identifiers, but I think in what I suspect you discussed yesterday, we do have to contend with the issues of information from that database being cross-classified, aggregate data being aggregated to the point so that you have small cell size, and that has the potential for in a sense, compromising this.

Mr. Gellman: Could you talk a little bit about the fields of information that you do have, so we get a better sense of that?

Dr. Edwards: The data content that we require our registries to provide to us include as I said, a study I.D. It contains socio-demographic information such as gender, age at diagnosis. One of the points that is now becoming a bit more controversial has to do with a geographic identification. So those are just some basic kinds of socio-demographic kinds of information to identify the case, but most of it -

Mr. Gellman: You said age. Do you have exact birth date?

Dr. Edwards: Month and year. We tend to convert that to age. We present it mostly by five year aggregates.

Mr. Gellman: What do you do for geographic identifiers? ZIP code?

Dr. Edwards: Each registry manages that slightly differently, but most registries use some geocoding process to take the address, which they actually collect, convert it into some geocoded location. At the moment, we have been having information provided to us at the county level. There is some interest in that information being made available at the ZIP code

level, or at the census tract level. As we move to those finer and finer levels, we end up having some concern raised about how much of that information is provided to us; how much of it would go on a public use file. In general, we have always had county-level data in part because we have often been asked to generate statistics at the general level. So we actually take the data on the cases, we obtain the populations from which they came, obtaining inter-censusal and projected population data from the Census Bureau for these geographic units that are within the SEER program, and calculate a cancer rate.

Mr. Gellman: You said that some of the registries collect addresses?

Dr. Edwards: All of the registries collect addresses.

Mr. Gellman: Do they keep the addresses, or do they just get converted into codes and then discarded?

Dr. Edwards: They collect it. How they manage that within a state is really within the state purview. It is not anything that we require, nor do we have any particular policy or procedure on that. I think most of them actually keep it, retain it, and it is certainly the confidential information. Part of the issue we face in cancer registries, this is a population-based cancer registry. We report the data on cases, that is individuals may be in the database multiple times, because a cancer can occur multiple times in an individual. The state-based registries or regional registries actually have to consolidate information from multiple reporting sources, whether they be hospitals, pathology labs, radiology facilities, clinical practices. So their requirement is to consolidate information on the same cancer case, and to try to resolve whether the information relates to that particular episode, that diagnosis. Is it a recurrence, or is it a second primary? In order to do that, they actually maintain a fair amount of what you would say is identified data. It is used from an administrative management standpoint, which we never see.

Mr. Gellman: You said that a case number is used.

Dr. Edwards: Study I.D. That is arbitrarily assigned. There is nothing unique about it.

Mr. Gellman: So the linking that is required is done without names, without social security numbers?

Dr. Edwards: They do that. We never see it.

Mr. Gellman: When you say "they do it," you mean?

Dr. Edwards: They, the state registries.

Mr. Gellman: They have the information?

Dr. Edwards: They do.

Mr. Gellman: But it doesn't flow up to you?

Dr. Edwards: It never flows up to us -- well, the only time any staff member would be aware of that information would be when - part of our program is to look at data that has ranched in quality of the data, the coverage, completeness, the accuracy of the information. As part of the SEER program we have always had a quality control activity which involves field studies, to go out and to actually audit records. We do that through each of the state regional-based registries. We have our staff who would go. Actually at the moment it's the design for how we conduct our quality control programs, which whether it be case finding, re-abstracting, or coding of the elements is planned in conjunction with our program directors of the registries.

It focuses on issues we think that relate to quality or coverage. So we actually do have staff in the National Cancer Institute, and staff through a contractor quality control unit who do go and access medical records that do contain confidential information on-site as part of case findings. Those individuals are required to follow all the rules of whatever is required of these medical facilities. That has actually been of concern at times, about the role of our quality control unit in actually having access to that information. That information is not removed from its source, but it is merely to be cross-checked, to make sure that the cases that are reported to us, that we have all the cases, and that the information is accurate.

Mr. Gellman: How are records linked across states?

Dr. Edwards: Each state has to negotiate what I call the reciprocity agreements. I'm not quite sure about your question about how are data linked across states. We are a geographically defined, population-based registry. The requirement is that data on all cases among those residents be reported. Many

of those residents may go to hospitals across the river, across the border. They may be retirees who go far away. That is an issue. Most states engaged initiate and engage in discussions with those places that have information that would be where their populations may be in order to acquire information that pertains to any cancer that has been diagnosed.

Those agreements are negotiated and are implemented at the state level under their authority. We do not become involved in any of those discussions. That actually is an important issue that relates to quality coverage and the aggregation of data on residents in those geographic regions. The data that we collect -- we talked about the geographic, the address question. Our primary purpose is on information that characterizes the cancer, the extent of the disease, the stage. We also collect a fair amount of information on treatment, though we focus on the first course of treatment, and mostly focus on modality.

We also are very interested, because we collect information on what happens to the patient, mostly survival status. We then actually track to find out whether the individuals are alive or not. We do not collect information on recurrence unless it would be through some special study. This kind of activity often is characteristic of the information collected in hospital-based registries. The state-based registries who have statutory authority now -- earlier some of them had administrative, public health authority for requiring that cancer be a reportable disease, and for if you will, running the cancer registry. We have not become involved in any way in what that authority is. Our requirement is that we have quality data, that it meets our audit criteria, and that information be provided to us in a certain time frame.

We then at the NCI, aggregate that information. We publish on it. We publish aggregate data, and we make public use files available for others who would want to use it. Our office provides an inquiries function. We get probably over 2,000 requests a year for information ranging from if you will, one piece of information, to our entire published report. We annually publish data on cancer incidence, survival, and mortality, but we use mortality data from the National Center for Health Statistics, and we present that for both our SEER geographic regions and for the nation. SEER only covers 10-14 percent of the total U.S.

So we have been in business since the early 1970s, and our data is cumulative. That is, all patients are registered and reported to us that have been diagnosed in calendar years 1973 for the areas that we cover. We augment that database each year, and so every year it is a cumulating cohort of information, which is then updated and maintained and reported to us, and then we analyze it and make that information available. We are making every effort we can to make that data more available. The more we make it available I think in some sense, the more we have to be concerned about who is using the information, and how the information is being used. For example, when we distribute our public use tapes, at the moment we do require that those who use it understand concerns that we have about tabulating data to generate small cell size.

Our public use file contains data on it that we feel is quality information, that is, if through our data collection quality control process we believe we have information that does not meet our quality standards, that information does not go on to a public use tape. The piece that I have not really spoken about has to do in our surveillance program with sort of what goes around the program to make it what it is. First of all, a lot of the registry data is being used for many purposes. Our first purpose was to try to track the national cancer burden; who gets its, what happens to them.

We have then utilized this registry-based program, either as a group, or other programs that go on in other states that are not part of SEER for epidemiologic research, or for tracking other measures that are not part of the central data that we require to be collected. We do that through sampling, through special studies, many of which if they go beyond what is mandated to be reported through the state registry, the legislation and the regulations would require IRB approval. So that is determined as to the nature of the information, whether it is validating existing information or whether it is collecting new information that is not necessarily reportable.

So we have used SEER over these years as a

source of identifying cases for special studies, to address issues of etiology, patterns of care, more detail on quality of life, more detail on actually the whole spectrum of care. We have a major study right now that is looking at prostate cancer. We call it health-related quality of care. It involves telephone interviews. Again, that would require a special IRB approval. It is being conducted by six of our ten registries. It is managed by an advisory committee, and a number of investigators. So we use the SEER data as a springboard for a fair amount of surveillance research. We augment what is sort of the routine data elements in the routine reporting. When we do that, that then puts us into the research arena, and requires all the policies and procedures that go along with that.

Mr. Gellman: Why aren't all the states involved in SEER?

Dr. Edwards: Well, cancer registries -- it's a hard question. We actually oversample and overrepresent some of the special populations. When we started the SEER program there were not that many states that had registries that had been involved in it with quite the length of time, nor had met some of the quality control completeness reporting. So there was an issue of either legislation was not in place. The activities of the registry that was there, there was not sort of an historical track record. We have had a few registries that have been in, and are out over time, but mostly there was not really I think a requirement that all states have a registry. There are states that have had registries, that were not included. This had to do with trying to acquire information on populations so that we had coverage of a broad spectrum of populations, and also in the sense that we oversample special populations, because of the small numbers. So it was actually based on what was the track record and the sort of experience with cancer registration.

Dr. Schwartz: Can you tell me who the users are of your information? Who requests information from you regularly, and are there any groups that you have turned down?

Dr. Edwards: Well, first of all let me just point out why not registries in every state. By referring to the CDC program, in 1992, there was legislation that required that the Centers for Disease Control -- well,

the secretary that was then delegated to the CDC, they institute a national program of cancer registries to either help develop registries in every state, or enhance existing ones. So Mary could actually speak to that. It's recent, and there has actually been a major change in the landscape in terms of legislation, the development of registries, and that activity there.

Who are our users? It really is quite broad. We, in our program, actually do not field many questions from patient advocates. Occasionally they come to us. Most of those questions go to our Office of Cancer Communications. Most of our requests range from health professionals, health researchers. We do acquire a fair number of requests from those who are trying to gauge the market. They are perhaps trying to decide is it time to move into an arena in terms of a drug. How many cases of this type? Sometimes we can answer that question, and sometimes we cannot, because they may ask a very focused question on a certain disease site with histology.

So we have a wide range of users from the public and private sector. Again, we have a lot of users from the public sector, but most of those are fielded through our cancer information service, that then accesses the published data. So our office does not really see that. Most are really health professionals and researchers that our office directly responds to. That is, of course, in addition to other federal agencies and the Congress and our senior directors, people who may be giving talks and want to have some specific information.

Mr. Fanning: Can you go into a little more about the CDC and their program, or maybe Scott could shed some light on this? I just want to get to the coverage thing.

Dr. Edwards: In 1992, the legislation was imposed in part for sort of -- actually, I think it got started with an interest in breast cancer, and the need for data that was available in regions and states other than what SEER covered. They started funding the program in 1994. I think that you more or less cover about 40-some odd states. There is funding to over 40 states now?

Dr. Wetterhall: Yes, currently we support 45 states, the District of Columbia, and three territories. The intent is to have nationwide coverage. The model

is very similar to what Brenda has been describing insofar as these functions are devolved to the state, so our role is in providing technical assistance, financial support, assistance in drafting model legislation, assistance in assuring quality.

Dr. Edwards: I think there is probably one other thing that I haven't made known to this group. Cancer registries, again, the first state that had one was Connecticut. That's in the 1930s. Hospitals have been involved with cancer registries probably since the very beginning of time, but certainly in the thirties, many run by pathologists or by clinicians. Cancer registries really builds upon the hospital-based registry. That is changing now, because of the way health care is changing. So that really has meant that the problem of maintaining the registries at the state and local level has changed.

Dr. Harding: I had a squamous cell cut off. I had cancer.

Dr. Edwards: You are not in our registry.

Dr. Harding: I just wondered if I was in the registry and didn't know it. Where is the consent started? At hospitals who have their registries, do they have consent when a person signs into the hospital? I'm just asking; I don't know.

Dr. Edwards: No, mostly not.

Dr. Harding: There is no consent in general?

Dr. Edwards: No, because cancer -- there are statutory requirements in most states for cancer to be a reportable disease. As it comes to the state level, there is no consent. At the moment. Nor as far as I know in terms of it being maintained in a hospital-based registry. Again, the intent there is the way the registry has been used at the hospital level is through -- again, you can hear more about it, and perhaps you know as much or more about it than I might from a personal level. The cancer registry actually has been used to track within the hospital, what is the patient load, what is happening. That information is often used in cancer conferences by the clinicians that are actually providing the care. That information is reviewed, discussed. Again, the intent is to try to improve patient management, and inform the group in the aggregate.

Ms. Regan: How specific do those state statutes tend to be in terms of the amount of the information that they ask to be reported? Is it just the disease be reported, if you are looking at questions of quality of care and quality of life?

Dr. Edwards: My understanding -- and again, I think the CDC, who has actually looked at almost all the legislation. As you have heard, one of the purposes of the NPCR was to talk about model legislation. My understanding is that it varies in terms of which cancers are reported. There is even an issue whether in situ is reported, that is, survival as an incidence-based only. So the specificity, as far as I know, varies. Now with the NPCR -- National Program of Cancer Registries -- the legislation actually identified some data elements. So I think the intent with those that have been developing or expanding registries through that program, the intent is to be a bit more uniform.

Mr. Van Amburg: I asked Bob to follow you, because for many years in my career I operated the cancer registry, along with some other registries. Actually the model for the cancer registry pretty much follows, at least in Michigan, the operation of the other large registries. Michigan's cancer registry went back to 1985, which preceded the national legislation on state cancer registries. One of the SEER registries is the Detroit metro region, so we already had 50 percent of our population covered in a registry before we actually had legislation. The legislation was spurred primarily by concerns in the environment in the 1980s. As I'm sure you realize, Michigan is completely surrounded by water almost, except for to the south. There was concern about the waters. We also had a problem in 1974, a fire retardant in cattle called PBB, polybrominated biphenyls.

So there was interest in the environment, and we did get statutory legislation for the reporting of cancer in 1985. That legislation has been kind of a model for the rest of our registry legislation that was enacted since that time. It basically requires the reporting of certain cancers, excluding skin cancers. It leaves pretty much the rest of what is reported up to administrative rule, which is actually the way you want to design something, so you have some flexibility to add items later on, or take items away that prove not to be valuable. In fact, the national legislation and the

requirements for new items, it was very easy for Michigan to add those items to its data set, because of the administrative rule process, by not having to go back through and change legislation. That legislation also specified access to the data at various levels, and I'll get into that in a minute.

Brenda did not indicate, I don't think, that the SEER registries for the most part, are active case ascertainment. At least in Detroit there are, where people go out and abstract the records from physicians and hospitals and laboratories. Most state registries that are not SEER supported are passive registries in which physicians, hospitals, laboratories report cancer cases to the state. Now clearly the active process is much more expensive, and it gives better results, and more complete results. The passive reporting works very well if you have a medical community that is interested in the registry, and you provide them with sufficient feedback to motivate their interest.

This registry was in operation since 1985. We now have well over 275,000 cases of cancer in the registry. We do collect identifiable data. We do keep it. There is very good reason for that. In Michigan we get about 45,000 cases of cancer reported per year. Of those, about 24,000 end up being unique cases. We have that much duplicate reporting which has to be weeded out of the registry. You really want that duplicate reporting, because at least in cancer, unlike some other things, people do shop around for diagnoses. They go to different physicians. They go to different institutions. They may go up to Mayo. They may go to M.D. Anderson in Texas. So we get multiple reports for people, and you need that identifiable information, at least on the front end, to be able to start matching these up, and eliminating the duplicate reports, then ascertaining if you have different information on cancers, which one of the diagnoses you want to keep. It is not a simple job.

Registries are really not very useful in their initial stages of development. It takes a number of years to get a database that is of sufficient size and quality to be able to really begin to publish data and start using it. With ten years of data in the cancer registry, we're just beginning to be able to look at some trends.

Let me mention one other registry while I'm thinking of that issue. The Michigan end stage renal disease registry, which had 20 years of operation and then was de-funded. It had just become very, very useful, because we could start looking at the long-term trends in end stage renal disease and transplants, and matching it to cancer registries to look at cancer incidence for who were on immunosuppressants and what have you, but it was de-funded.

As far as access to the data is concerned -- and I think a lot of states operate this way; I can't speak for all of them -- but we use four types of level of access. One is we publish data which are clearly non-identifiable, and in tabular form. We produce public use tapes, and we spent a lot of time agonizing over how to do this. It is not simple. We ended up with five different types of sets of tapes that people might want to use, so that we didn't include all the data elements in any one tape. We were pretty sure, but never absolutely certain that you couldn't identify an individual. Some people wanted month of the person's birth, so if we did that, we would eliminate something else geographically, so that you couldn't try to identify people. It was an interesting exercise. We do the same thing on the birth defects registry and other registries like that.

If those two methods don't work, we would try to convince the person that we would do the tabulations for them. This worked very well until the advent of personal computers and the explosion of the software to analyze data on personal computers. Most people did not want to work with large data sets. It was much easier for us to do it. They wouldn't make mistakes. We wouldn't be getting calls all the time about what something meant, and we were pretty sure of the data.

If that didn't work, then we had in the legislation, an advisory committee of what we call a technical scientific advisory committee to the cancer registry, and we use it for the other registries as well. It would review applications for access to data that might reveal the identity, or required the identity of an individual for an epidemiologic follow-up exam. I think Massachusetts has a similar approach. This committee we had appointed from the research and medical community, primarily from the provider community. One of the requirements before this scientific

committee would even look at the application was it had to be reviewed by an IRB first. They were not in the issue of ethics or what have you. They were interested in the scientific validity of the data.

The legislation was set up specifically so that the registry would not be misused or used haphazardly by people that are purporting to do scientific investigations without merit. This committee would review applications all the way from we get applications from Harvard, and we get applications from U of M and Ford Motor Company and Dow Chemical and students. It would review these and either make recommendations to modify the design, accept the design, or reject the design. The committee did reject the designs that they determined that based on the type of study done, and the number of cases that knew were in the registry, they could not ascertain the information they really wanted to do. We have turned down requests for data for lists of people with certain types of cancers. You will get these. Some of these are coming from well meaning groups, advocacy groups, support groups. These get turned down. There is no administrative use of the data required allowed by the legislation. So far, as far as I know -- and you never really know I don't think -- there has not been any misuse of the data file that we have let out for any research.

One of things that we do with that registry that was not mentioned by Brenda, but very briefly, is that we add to that registry from another registry, which is the mortality registry. There are cancerous types of diagnoses that are not determined essentially until death, and so we search out from the mortality registry, all the cancer deaths. Then we have to follow back to the physician to get the data incidence of the cancer, because that is the key item, beside the diagnosis, is the incidence tape, the first diagnosis tape. So that is a case where we have added registries together essentially, and then we match the mortality registry, the cancer registry to ascertain who has died and who is still living.

I want to mention one other thing that Brenda touched on, and I think it is a very critical problem for the future, and that is the GIS systems **-- geographic information systems --** and the mapping of data,

which seems to be in great favor right now. This creates a whole different level of confidentiality and privacy, because while you are not revealing the name of an individual, how you close you map that out can reveal the identity of an individual. If you go down the block at the census tract level, you can have some real problems. So we have been really working to try to ascertain how to best be able to display data with the geographic mapping techniques, and to not divulge confidentiality, because if you are looking for clusters, they may occur in small areas. You know that is going to create a problem.

Brenda mentioned one other issue, and I want to take the reverse of it. She was concerned about small cell size. I am concerned about large cell size. If you have a population of people in an area, let's say a county, and if you have a population let's say of Arab Americans or Hispanics or something in a population that is not large, but not small, let's a couple hundred people, and your tabulations show that 85 percent of those have that diagnosis, you have revealed identity essentially, because the probability of anyone in that community having that diagnosis is 85 percent. So you have a reverse problem with small numbers as well, and we have been trying to work with that issue as well. It is a very interesting area, especially when you have enclaves of population groups.

****

Web Site(s) of Interest for this Edition
There are four sites of interest. The first is a Powerpoint presentation of 29 slides shown at the recent May 15 University Consortium for Geographic Information Science (UCGIS) Congressional Breakfast at http://urban.rutgers.edu/ucgis. This provides a good visual overview of GIS, adapted to a congressional audience. Be sure to see slide 18 if you wish to add hemlock wooly adelgid to your vocabulary. Senator Susan Collins remarks are there also but, unfortunately, Senator Domenici's well received comments were made without written text.

The second is Bill Bowen's "Digital Atlas of Washington, D.C. and Vicinity" which is his latest addition to the collection of urban, socioeconomic and

demographic maps of major U.S. cities. It currently displays 128 color maps of the nation's capital and can be viewed at http://130.166.124.2/dcpg1.html. Many of these maps show census statistics by census tract.

The third site will take you to the recently established National Institute of Justice (NIJ) Crime Mapping Research Center(CMRC). The goal of the CMRC is to promote research, evaluation, development, and dissemination of GIS technology. The CMRC may be viewed at http://www.nlectc.org/cmrc/. Some related editorial comments are presented in the Final Thoughts section.

The fourth site "Just Another Medical Geography Page" is located at http://www.geocities.com/Tokyo/Flats/7335/medical_geography.htm and has been updated as of April 10, 1998. This website is devoted specifically to the field of medical geography and contains links on the geography of infectious and noninfectious disease, the geography of health and health care, GIS and remote sensing applications in health, a bulletin board, chat room, conference calendar... and literally hundreds of links to researcher homepages, electronic discussion groups, lesson plans, published studies, shareware and freeware, and more. This month, two new sections have been added: a directory of online health atlases, and the first edition (to arrive shortly) of Medical Geography Digest.

****

On the Lighter Side of GIS

Three roommates slept through their midterm geostatistics exam on Monday morning. Since they had returned together by car from the same hometown late Sunday evening, they decided on a great little falsehood. The three met with the instructor Monday afternoon and told him that an ill-timed flat tire had delayed their arrival until noon. The instructor, while somewhat skeptical, agreed to give them a makeup exam on Tuesday.

When they arrived the instructor issued them the same makeup exam and ushered each to a different classroom. The first student sat down and noticed immediately the instructions indicated that the exam would be divided into Parts I and II weighted 10% and 90% respectively. Thinking nothing of this disparity, he proceeded to answer the questions in Part I. These he found rather easy and moved confidently to Part II on the next page. Suddenly his eyes grew large and his face paled. Part II consisted of one short and pointed question......."Which tire was it?" [Editor: modified slightly from Gary C. Ramseyer's First Internet Gallery of Statistics Jokes at http://www.ilstu.edu/~ gcramsey/Gallery.html]

****

Final Thought(s): Exciting GIS Future in Criminal Justice

Attending and participating in Towson University's annual GIS conference (directed by Jay Morgan, Ph.D., Department of Geography and Environmental Planning) is always special to me and this year was no exception. Nancy La Vigne, Ph.D. and Director, Crime Mapping Research Center (CMRC), National Institute of Justice (NIJ), gave an insightful plenary presentation entitled "The Use of GIS for Criminal Justice Research and Practice." GIS development in criminal justice is emerging as an important tool for improving crime prevention which, from any perspective, is an integral part of public health. For example, in New York City, a COMPSTAT model successfully blends computers and statistics in a GIS environment to aid in the timely response to crime in real time and to the detection of crime clustering and patterns. This use of GIS has been credited as a key element in the reduction of New York City crime.

There is other emerging evidence to support the versatility and innovative use of GIS in the criminal justice system. One, some police departments are starting to use the Web for posting crime statistics and maps. Two,

some prison cell assignments are being made with greater sensitivity to prisoner origins where prior association with areas of infectious diseases such as TB and AIDS parlay into the assignment equation. Even the management of prisoner flows e.g., through time and space, are being tested with GIS. Third, orthophotoquads of crime areas have been used by U.S. attorneys to help validate plausibility tests related to the timing and navigation of criminal acts. Fourth, new uses of GPS with electronic monitoring are being tested including community notification (Megan's Law) of sex offenders. Fifth, GIS is being tested in the assignment of risk and route for improved community law enforcement through spatial statistical correlation. Lastly, predictive modeling with GIS through neural network analysis is being tested to show not just where hot spots occur but where, in the case of drug markets, they can be expected to become established.

The CMRC funds over 30 grants for crime mapping projects including the development of analytic methods software for crime analysis (see Web site above). Dr. La Vigne concluded by mentioning that many police departments are just now investing (still few users) in GIS and that the First National Crime Mapping conference was conducted last year. [Contact: Dr. La Vigne at voice (202) 616-4531 or email lavigne@ojp.usdoj.gov]

August 17-20, 1998 "GIS in Public Health" Conference