



August 31, 2009

Maryann Catalano, Senior Vice President
New York City Economic Development Corporation
110 William St, 6th Floor
New York, NY 10038

Re: Public Access to New York City Government Data & Related Software Application
Competition

Dear Ms. Catalano:

Please consider this letter as our response to the above-referenced Request for Expressions of Interest (RFEI), otherwise known as the City's "NYC Data RFEI" related to the City's upcoming "Big Apps" competition.

Our experience with New York City agency data

I direct the CUNY Mapping Service, a project of the Center for Urban Research at the City University of New York Graduate Center. The Mapping Service engages with foundations, public agencies, businesses, nonprofits, CUNY researchers, and others to use spatial information and analysis techniques to develop and execute applied research projects. Our expertise lies in the use of geographic information systems (GIS) to understand, visualize, and analyze data sets for a variety of urban planning issues including demographic change, land use trends, social service availability, educational needs, public health, and environmental quality. More information about our work is available at www.urbanresearch.org.

The Mapping Service (and the Center for Urban Research) has extensive experience with City data. We have worked either in partnership with several New York City agencies or on a contractual basis to analyze, visualize, and provide strategic access to dozens of City data sets.

For example, we maintain the Open Accessible Space Information System (OASIS) at www.oasisnyc.net which integrates several dozen layers of mapped data – not only from City agencies but from all levels of government as well as private sources in the five boroughs and in northern New Jersey – to visualize the nexus between community greening resources and broader urban planning issues. The OASIS website was developed in partnership with the USDA Forest Service and a coalition of open space organizations including the NYC Department of Parks and Recreation. It provides easy access to view and map the wide variety of open space resources in the city, and how these resources are related to and interconnected with land use patterns, transit, schools, housing, and more. It is free for anyone to use.

In 2005, CUR partnered with the City's Department of Finance to evaluate the effectiveness of handheld devices enhanced with GIS maps to improve the efficiency and accuracy of data collection for property valuation and tax assessment purposes. The work involved converting the Department's Computer Assisted Mass Appraisal (CAMA) data into a format that was readable by handheld devices and also accessible via a wireless connection. It also involved mapping the data using GIS layers from the City's Department of Information Technology and Telecommunications (DoITT).

Over the years we have been actively engaged in educational efforts around the importance of removing barriers to data access in ways that benefit the City, the general public, and key constituencies. The CUNY Data Service at the Center for Urban Research has for years hosted the annual conference of the New York Area Data Council in partnership with numerous city, state, and federal agencies, businesses, nonprofits, and individuals. In 1999 the Data Council issued a paper which I co-authored titled *Making the Best Use of Government Data*. I have included the executive summary with this letter. Even 10 years ago, the paper laid out a set of guidelines that still describes an appropriate and forward-thinking context as you plan for the Big Apps competition.

Benefits of making city data widely accessible

Proactively sharing data sets and making them available electronically (online or otherwise) in multiple formats without burdensome license restrictions or fees provides several benefits. It helps create a level playing field for residents and others to participate meaningfully in local decision-making. It fosters transparency in government and a sense of openness. It provides benefits to the City directly, such as:

- **Cost savings:** staff time for legal review and responding to ad hoc data requests will be minimized or eliminated;
- **Improved data quality:** agencies will receive feedback, corrections, and/or updates from a much broader constituency, helping agencies meet their regulatory requirements;
- **Economic benefits:** startups and other companies that are developing apps that rely on City data will generate tax revenue, jobs, and business activity. Conversely, a lack of access to City data hinders New York's competitiveness with other municipalities that provide easy access to their data.

Public access to municipal data can also streamline the workflow for public employees at all levels of government – such as the Metropolitan Transit Authority using property-level records to evaluate the impacts of the East Side Access project, or the state's Department of Transportation integrating data about the City's roads into an assessment of statewide bridge inspections, or the US Census Bureau seeking to improve its address lists based on the latest tax parcel addresses from the City's Department of Finance.

What data sets should be made available?

The City should take the broadest view regarding data access, following the spirit and letter of the state's Freedom of Information Law (FOIL). Any and all databases, lists, and other data elements maintained by City agencies and staff should be considered fully in the public domain unless they meet one of FOIL's limited exemptions.

In this electronic era, virtually everything City government does involves computerized data. Even in 2001, an internal inventory for DoITT identified more than 100 electronic data systems in use by City agencies at that time, each one for a particular government function. Some of these have certainly been changed, become obsolete, or subsumed into more recent initiatives such as 311. But the list includes myriad data items that would be of great help to people interested in evaluating City services, monitoring agency programs, and simply trying to better understand how their government works. Virtually all of these databases should be opened up publicly, either for use in apps or otherwise.

Many City agencies have greatly improved their data access policies and practices, such as the Department of City Planning's Bytes of the Big Apple data. For example, City Planning used to charge a fee for obtaining a GIS database of the City's street grid, which is now available for free download without license restriction. But City Planning still requires fees and licenses for other electronic data sets, such as its tax parcel databases known as "PLUTO" and "MapPLUTO".

These tax parcel files can be extremely helpful for property development activities and community planning efforts. At a time when economic recovery and rebuilding has such a high priority in New York, easy access to data and tools that can help more organizations and individuals participate in these activities can only help both the economic recovery of the City and the quality of our built environment. The City would help achieve these goals by reducing or eliminating fees for PLUTO/MapPLUTO, as City Planning already has done for its other GIS data.

In general, the City needs to take a hard look at what purpose and value (if any) is served by limiting access rather than providing open access with few or no restrictions. A growing number of local governments in New York and across the country are realizing that the benefits of open access outweigh the perceived risks.

In my experience, the City needs to include four critical components with any data they make available. These are:

1. **Geography** – if the is already maintained in a GIS it should be provided in standardized formats such as shapefile, KML, or GML. If that data is not in GIS format but includes information about geography (street addresses or ZIP Codes, for example), this information should also be provided in a standardized format so the data can be converted to and/or integrated with other GIS projects.
2. **Attributes** – information about each record that describes it in some detail (in other words, not simply the name of a licensee but the license category, regulatory program, etc).
3. **Metadata** – elements such as:
 - a. The data's original purpose (why it was collected/created in the first place – this helps potential users know if their intended usage is appropriate or not);
 - b. Its vintage (when it was created);
 - c. Its level of precision (usually pertaining to geographic information); and
 - d. How often it is modified/updated.
4. **Common identifiers** – so people can link data element to related information. Examples include:

- a. school ID to link to the state Education Department school report card;
- b. building identification number (BIN) to link to Department of Building data and/or Housing Preservation and Development enforcement information;
- c. a borough-block-lot (BBL) tax parcel ID to link to assessment, ownership, and other property-level data;
- d. a standard ID from the Department of Parks and Recreation to link parks to budget information.

Below is a list of specific data sets that either have not been made publicly accessible or are only available on a per-request basis. We believe these should be included not only in the Big Apps competition but should be made publicly available electronically without license restriction or fees. The data should be freely downloadable via public websites (as well as available in more traditional formats including CDROM or printed lists). The examples are grouped by category, in no particular order:

Licensees and contractors

- According to the Department of Consumer Affairs website (www.nyc.gov/html/dca/html/licenses/licenses.shtml), “DCA licenses more than 60,000 small businesses Citywide in 55 different categories. In order to operate legally in New York City, many businesses must obtain a license from DCA.” These include billiard rooms, cigarette retailers, parking lots, newsstands, sidewalk cafes, and street fair vendors. The list of licensees in each category should be made available along with information about each one. Applications using this information might include maps showing densities over time in relation to land use, zoning, transit routes, and public services, as well as mobile applications identifying these locations dynamically.
- Similarly, the Department of Health and Mental Hygiene issues licenses in 11 major categories (www.nyc.gov/html/doh/html/permit/license_permit.shtml) including child care providers, food vendors, facilities that use radioactive materials, and funeral homes. This data should be made available.
- Vendex data on public contractors. This information is extensive and complex, but elements of it would be interesting and useful for application developers and the general public, such as contracts given to businesses registered in the city vs. outside the city, types of contracts by vendor, vendors by agency, etc.

Environmental quality

- The Department of Environmental Protection (DEP) collects data on myriad facilities and sites that have may environmental impacts such as air pollution sources, industrial wastewater dischargers, combined sewer outfall (CSO) drainage pipe locations, brownfields, and hazardous materials complaints (historical and current). City, state, and federal regulations require DEP to maintain a wealth of data on each of these, including location, type of environmental impact, amount of material such as hazardous waste or wastewater discharge, history of environmental impact, etc. Though state and federal environmental agencies also collect and maintain similar data, DEP often has unique data not included in these state and federal files such as the City’s “industrial pretreatment program” or CSO locations. Yet these state and federal agencies now make most of their data available via public download. Once DEP makes its data easily accessible, this

information can be combined or integrated with state and federal sources to create a more complete picture and analysis of environmental impacts facing city residents, businesses, and visitors.

- The Department of Parks and Recreation maintains an inventory of properties and environmental resources including parks as well as Greenstreets, Forever Wild areas, recreational sites, and more. This is available by request, but should be publicly accessible.

Property data

- The Department of City Planning's MapPLUTO database showing the mapped patterns of property ownership, detailed land use types, and assessed value is available only by signing a license agreement and paying a fee of up to \$1,500 per year (whether you are requesting the data for a public agency, private company, or local community group). Either DCP's MapPLUTO system should be made publicly available without licenses or fees, or the individual elements of that system – such as the parcel boundaries in GIS format either from DCP or Department of Finance's new Digital Tax Map, DOF's Real Property Assessment Database, and the Major Property Owner files and Integrated Property Information System (IPIS) from Department of Citywide Administrative Services – should be made publicly available.
- The Department of Finance also has extensive data on historical property sales in the city. Recently DOF has provided free downloadable data for recent sales, but only back to 2003. The Department's internal data systems maintain this information at least back to the 1980s, and the full set of historical data should be made available.
- DOF's "Automated City Register Information System" (ACRIS) provides a wealth of data on property transfers, deeds, etc. The online system to access this data on a property-by-property basis is impressive. But "batch" versions of the data are only available through specific requests and/or fee-based subscriptions and in a cumbersome, non-relational ASCII format. The underlying ACRIS data should be publicly available so application developers and others can analyze and monitor this information, at least historically if not in real-time.

Buildings data

- The Department of Buildings Information System (BIS) provides a wealth of information of keen interest to the public and to anyone hoping to understand development trends, building safety, and how the City interfaces with the real estate industry in general. BIS includes information on building applications, certificates of occupancy, inspections, (final construction, plumbing, electrical, and elevator), complaints, violations, safety reports, licensing data for contractors, and more. This should be publicly available beyond the building- or property-specific searches via DOB's website.
- The New York City Loft Board maintains data on all "interim multiple dwellings" in the city. Although there are relatively few of these – 318 as of August 2009 (www.nyc.gov/html/loft/html/buildings/buildings.shtml) – loft living is a subject of interest to residents, urban planners, and others. The Loft Board provides a list of IMDs on its website, but only in PDF format (www.nyc.gov/html/loft/downloads/pdf/imd_buildings.pdf) and the list omits other

information such as violations. The list and related data should be publicly available in electronic format.

- Similar to DOB, the Department of Housing Preservation and Development (HPD) provides a searchable online system (HPD Online) to identify housing violations and enforcement actions on a property-by-property basis. But the underlying data should be made fully available so individuals, researchers, and others can understand the patterns, trends, and other information about these issues.
- The Landmarks Preservation Commission maintains data on individual landmarks, interior landmarks, and historic districts throughout the city. It provides PDF maps of districts on its website, it used to provide downloadable GIS files of district boundaries, and it provides databases of landmarked properties on a per-request basis. This data should all be publicly accessible.

Education

- GIS files of school district boundaries are publicly available for free via the DCP Bytes of the Big Apple program. But the boundaries of school *zones* – the smaller areas within districts that generally provide the determining factor as to which elementary school your child can attend – are not accessible except via FOIL request. The zone boundaries are maintained by the Department of Education in GIS format and should be made as easily available as the district boundaries. These boundaries can change over time more frequently than districts, and historical versions of the boundaries – as well as RSS feeds of changes as they occur – should also be made available.

In addition to the actual data sets, notifications when each data set is posted and when it is updated should be available via RSS feeds. Some agencies have begun to implement this approach, such as City Planning and Finance. But like other cities (Washington, DC in particular), anyone should be able to subscribe to data updates based on the data categories of their choice. Or emails could be provided, a la New York City Transit's "Subway Service Advisory Email Notification Program" at <http://advisory.mtanyc.info/index.html>. Another example is the US Environmental Protection Agency sending emails whenever its Geospatial Data Access Project data layers have been updated. These automated updates will be especially important as the City steadily expands the number of publicly available data sets from multiple agencies.

It will also be important for agencies to plan for feedback mechanisms. The City should look favorably on applications that not only use City data but also provide opportunities for the public to comment on the information and/or provide detailed updates to each agency that created the original data. For example, the OASIS website receives emails from property owners in the city who believe that the property data displayed via OASIS (which was obtained from MapPLUTO and originally from the Department of Finance) is incorrect or outdated. But the city has not provided an easy, automated mechanism for submitting this information so it will get addressed by the appropriate agency(ies). Once data is opened up and applications are developed with widespread exposure, the City should not only expect but should encourage a substantial increase in public commentary related to the accuracy of this data, and plan accordingly to integrate the commentary back into the original data systems.

Finally, I urge the New York City Economic Development Corporation as the agency receiving the RFEI comments to publicize all the RFEI submissions. We are part of a growing community of individuals and organizations eager to compare notes and work in collaboration on the best ways to improve the quality of life in New York through improved public access to City data. You can help facilitate that effort by enabling us – and the public at large – to view each other's submissions.

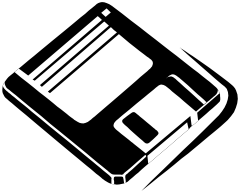
Thank you for the opportunity to submit these comments.

Sincerely,

Steven Romalewski
Director, CUNY Mapping Service
c/o Center for Urban Research
The Graduate Center / CUNY
365 Fifth Ave., Room 6202
New York, NY 10016

Phone: 212-817-2033

Email: sromalewski@gc.cuny.edu



New York Area DATA COUNCIL

July, 1999

In the information age, data is gold--and many government agencies are letting it drain away, depriving even colleagues in their own unit of government of valuable information, wasting taxes needed to generate the data anew. This paper, *Making the Best Use of Government Data*, strongly encourages an atmosphere of sharing data in forms that other organizations can use. It addresses difficult problems of privacy, secrecy and cost, recommending more public discussion of those issues.

BOARD OF DIRECTORS

Paul Barnett
Jannette Barth
James Brown
Edward Casey
Jack Eichenbaum
Marta Fisch
Leonard Gaines
Jackie Gold
Michelle Holder
Shirley Jelks
Nathan Kantrowitz
Stephen McDevitt
Donald Menzi
Bruce Mesh
Yuko Nakanishi
Zvia Segal Naphtali
Michael Pferr
Yousuf Rahman
Timothy Ross
James Savitt
William Shore
Thomas Spitznas
Raymond Spungin

The paper includes valuable information about data quality and current availability and the opportunities and problems generated by more data sharing. It summarizes coordinating efforts of federal, regional, state and local government and the private sector.

Some of the recommended action must be initiated by elected officials. Other improvements can be carried out within government agencies by spreading a culture of data openness, assuring high quality data and making it available in forms usable outside the originating agency.

The authors hope the report will encourage discussion of these issues and result in more useful information available to government, business, personal decisions and serious research. We foresee a reduction in the cost of high-quality information and an expansion of protections against invasion of privacy.

We welcome responses to the recommendations. Please fax back written comments to Jack Eichenbaum (212) 669-2060 or email jaconet@aol.com. We would meet with interested officials to further the usefulness of our recommendations.

Below is our report: *Making the Best Use of Government Data..*

SPONSORS

- CUNY DATA SERVICE

- EMPIRE STATE
DEVELOPMENT

- NEW YORK CITY
OFFICE OF
MANAGEMENT AND
BUDGET

Executive Summary: Making the Best Use of Government Data

Data sharing among public agencies and between public agencies and private organizations is essential to take advantage of the information age efficiently -- to save costs, skilled labor and time in data collection and maintenance. To promote that practice, we strongly recommend that:

1. Executive offices of government at all levels should establish firm policies in favor of data sharing. They should increase public servants' awareness that data collected for their agency needs is likely to have value to other organizations performing different functions.
2. Public policy should promote cooperation among agencies, public and private, in data collection, maintenance and interchange. The New York State Geographic Information Systems (GIS) Cooperative and the New Jersey State Mapping Advisory Committee provide leadership in these issues.
3. State legislatures should foster the data sharing process with guidelines that promote (a) public acknowledgment of the data source and its ownership and (b) cooperative relations between data users and providers.
4. Regional data is difficult to gather where there is no regional government presence. Regional indicators assist planning and other important government decisions, and the collective data must come from somewhere. An organization such as the New York Metropolitan Transportation Council should be financed to collect data of economic and social interest at the tri-state regional scale.
5. Within the broad principle of making data widely and easily available, privacy and secrecy must be protected. Individuals should have the right to legal recourse when their privacy has been violated, including the right to correct false information. Legislatures should set policy and agencies should be held responsible for carrying out that policy.
6. Useful data can often be provided without violating secrecy and privacy if it is suitably aggregated or summarized. Government agencies that recognize the overall benefits of freely available data will take the extra steps necessary to provide it in a form that protects privacy and secrecy.
7. Agency heads and budget offices should recognize that data collection and maintenance are integral components of many public agency goals and should fund them consistently. Data that is not current often loses its value.
8. Public offices should assure professional standards in data handling. High quality dependable data free of wrongful disclosure requires that.
9. Data collecting agencies should provide a full description of data (metadata) with distribution. Metadata should detail data encoding, data gathering and sources of error.
10. The provision of data is useful only if prospective users know the data is available. Governments should establish and publicize available data -- using all forms of publication to accomplish this.
11. Data distribution format is important in facilitating its use, especially by organizations that don't have highly professional research capacity. Governments should provide information on paper as well as in electronic forms, laid out so it is easy to understand and analyze. Data clarity also requires eliminating irrelevant material.
12. Data users have a responsibility to data providers. They must acknowledge the source, sending to the data source research based on data provided, commenting on the accuracy and clarity of the data to the source, and, most importantly, using it with care to avoid misinterpretation.

13. Recognizing that there are costs to making data usable and widely available, users might legitimately be asked to pay the full marginal cost, especially if the user will profit from use of that data. However, government data should be available to the public free of charge where it is essential to assure responsive and responsible government.
14. Some issues should be clarified by further public debate:
 - ◆ Where market pricing may be appropriate
 - ◆ Individual right to privacy and recourse when violated

Additional conclusions are suggested in the full report. Some concerns were not resolved and are posed as questions inside the document. For other dilemmas, opposing views are aired.

Contributors

Jack Eichenbaum chairs the Public Data Access committee of NYADC, which produced this paper “*Making the Best Use of Government Data*”. Dr. Eichenbaum has a Ph.D. in urban geography and coordinates GISMO, New York City’s Geographic Information System user group. Dr. Eichenbaum works with property data as a City Assessor in the Appraisal Research Division of the New York City Department of Finance and has published several articles on the value of municipal data.

Cathe A. Sheehan, a technology and information strategist, is currently consulting to a major global corporate banking and capital investment service. She has significant work in commercial market research and economic statistics. Her career has emphasized considerable technology concerns for organizations ranging from the Federal Reserve Bank and U. S. Department of the Treasury, to a variety of commercial enterprises: including financial, trade and professional groups, manufacturing, and health care.

William Shore is a Senior Associate of the Institute of Public Administration and Secretary of the Regional Research Consortium, sponsored by the Institute, several prominent universities and the New York Metropolitan Transportation Council. He joined the Institute after 35 years with Regional Plan Association, where he was responsible for public information and pioneered in public participation projects. Previously, he was Publications Director of the American Society for Public Administration and editor of *Public Administration Review*.

Stephen McDevitt is a geographer with the U. S. Army Corps of Engineers, New York District, Planning Division, Environmental Assessment Branch. He also serves as the Geospatial Data and Systems Point of Contact for the district. As the Point Of Contact, he is responsible for arranging geo-spatial data sharing agreements with various federal, state and local government entities within the New York district borders. He previously worked for the New York district, Real Estate Division’s Cadastral Section, as a cartographer technician, scrivener, and physical scientist.

Steven Romalewski coordinates the Community Mapping Assistance Project (CMAP) of the New York Public Interest Research Group (NYPIRG). CMAP provides customized mapping services to non-profit and other organizations throughout New York State. Prior to CMAP, Mr. Romalewski had 15 years experience at NYPIRG accessing public data for community groups and others concerned about environmental health.

Other members of the New York Area Data Council’s Public Data Access Committee participated in initial discussions: **Bruce Mesh**, Bamaware Corporation; **Zvia Naphtali**, NYU Wagner School of Public Administration; **Marta Fisch**, CUNY Data Service; **Leonard Gaines**, Empire State Development Corporation; **Nathan Kantrowitz**, NYC Department of City Planning; **Shirley Jelks**, Metropolitan Transit Authority