Data Sharing Best Practices for Local Governments in the Chicago Region
November 2011

Purpose

The purpose of this document is to describe best practices for local governments wishing to make their data available to the public and other government entities. This is in accordance with action steps defined in CMAP's GO TO 2040 comprehensive regional plan, specifically those under the recommendation to "improve access to information." Among the action items associated with this recommendation are the following:

Define Best Practices for Data Sharing

By understanding how communities here and across the world address similar data challenges, our state and region can put these best practices to work for residents. CMAP will work with the State of Illinois, counties, municipalities, and other governmental bodies that possess data to implement technical improvements that facilitate open exchange of data.

Provide Technical Assistance

CMAP will help local governments post data online and will host a regional online warehouse to provide access to municipalities' information as an expansion of the Regional Indicators Project. CMAP will also seek opportunities to form partnerships around data-sharing pilot projects with other interested regional partners, such as transportation agencies, workforce development agencies, and municipal water suppliers to make this data more accessible.

(See http://www.cmap.illinois.gov/2040/efficient-governance .)

Background

Data sharing by government entities has been an important issue in the United States since the earliest days of the republic. The first United States Congress debated--and tabled--a proposal that would have allowed journalists to cover its deliberations. [1] The Freedom of information Act (FOIA) as we know it was enacted in 1966 over the objections of President Lyndon Johnson. State FOIA laws soon followed, and today every state has some form of FOIA statue. Most have open meetings laws as well.

Transparency vs. Inter-governmental Data Sharing

Government data sharing includes two distinct but related contexts:

- Inter-governmental data sharing
- Disseminating government information to the general public

Information sharing with the general public is often called ***transparency***, which also describes non-data contexts such as open meetings. Laws governing transparency are sometimes called ***sunshine laws*** or ***open records***

laws. For purposes of this document, ***data sharing*** will refer to the general public availability of data, documents and other information artifacts created by public entities. ***Transparency*** will refer to the broader context of open data, open meetings and public access to governmental processes. ***Inter-governmental data sharing*** will be used to describe data sharing that explicitly excludes the general public. Government data sharing can include:

- Administrative data such as meeting minutes, contracts and financial records
- Operational data such as service requests received, permits and licenses granted, code violations issued, crimes prosecuted, emergency responses, and so on
- Measurements such as traffic volume counts, results of public surveys, and so on

Importance of Inter-governmental Data Sharing

The importance of inter-governmental data sharing has been recognized for many years, and nowhere as explicitly as in criminal justice. The U.S. Department of Justice notes that "[A]s early as 1930, when the first interstate teletype connection went into service, law enforcement sought to exchange information as a potent weapon against crime." (http://it.ojp.gov/default.aspx?area=nationalInitiatives&page=1081)

The attacks of September 11, 2001 exposed the dangers inherent in government balkanization, especially within U.S. intelligence agencies. Experts argued that at least some of the attacks might have been prevented if law enforcement, intelligence and regulatory agencies had shared data among themselves. This criticism was not limited to federal agencies, since it became known that 9/11 ringleader Mohammed Atta had been stopped by local police for multiple traffic violations in Florida. A warrant was issued for his arrest after he skipped a court hearing, but the warrant was not picked up by other government agencies.

Support for data sharing has also been lively outside the Justice Department. For example:

Transportation – "Data allows for complete understanding of: the nature, causes, and injury outcomes of crashes; and the strategies and interventions that will reduce crashes and their consequences." (U.S. Department of Transportation, The Importance of Sharing Data, http://www-nrd.nhtsa.dot.gov/Pubs/810687.PDF)

Public health— "Information systems play a central role in developing an effective comprehensive approach to prevent, detect, respond to, and manage infectious disease outbreaks of plants, animals, and humans… However, access to some of these data sources and related search and reporting functionalities may be limited to the agencies that have developed such systems… reducing the effective use of infectious disease data in the national and global contexts." (National Science Foundation grantees, http://dgrc.org/dgo2004/disc/presentations/crisis/zeng.pdf)

Environment and natural resources – "Environmental data that affects important health issues also spurs a need for interagency coordination and efficient, timely exchange with regional and state health organizations." U.S. Environmental Protection Agency's National Environmental Information Exchange Network, http://fcw.com/articles/2003/09/29/epa-info-network-taking-shape.aspx)

Importance of Government Transparency

Transparency has long been considered essential to good governance and public accountability. James Madison wrote in an 1822 letter:

> "[A] popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or, perhaps, both. Knowledge will forever govern ignorance: And a people who mean to be their own Governors, must arm themselves with the power which knowledge gives." (U.S. Department of Justice, http://www.justice.gov/oip/foiapost/2008foiapost12.htm)

President Barack Obama wrote in 2009's Open Government Directive:

> "Transparency promotes accountability and provides information for citizens about what their Government is doing. Information maintained by the Federal Government is a national asset. My Administration will take appropriate action, consistent with law and policy, to disclose information rapidly in forms that the public can readily find and use. Executive departments and agencies should harness new technologies to put information about their operations and decisions online and readily available to the public. Executive departments and agencies should also solicit public feedback to identify information of greatest use to the public."

The American Recovery and Reinvestment Act (ARRA) was an important test case for the new transparency mandates. A blog post on recovery.gov in 2010 stated the benefits of ARRA transparency:

- The public knows what the government is doing with their tax dollars – for example, Recovery tax benefits are described in a simple chart with detailed narrative descriptions;
- Citizens can interact with government officials to influence decisions in a collaborative manner – descriptions of each award are displayed by the website's mapping capabilities giving the press and public the information to question Congress, the Administration, and state and local governments regarding the value of these efforts;
- With the information on projects and awards, the public can track the progress on the ground where funds are being spent – the awards in each local community can be identified using a zip code search;
- Information on opportunities is readily available – for example, there are two job search capabilities and links to information on available grants and contracts;
- Compliance and self-correcting behavior —all reports filed by those receiving funds are displayed and those who fail to report are named on Recovery.gov. In evaluating the relatively small number of award recipients who have not reported as mandated by the Act, it is obvious that the public nature of reporting and the open availability of the data do indeed foster compliance and self-correcting behavior.

The Illinois Freedom of Information Act describes transparency as follows:

> Pursuant to the fundamental philosophy of the American constitutional form of government, it is declared to be the public policy of the State of Illinois that all persons are entitled to full and complete information regarding the affairs of government and the official acts and policies of those who represent them as public officials and public employees consistent with the terms of this Act. Such access is necessary to enable the people to fulfill their duties of discussing public issues fully and freely, making informed political judgments and monitoring government to ensure that it is being conducted in the public interest.
>
> The General Assembly hereby declares that it is the public policy of the State of Illinois that access by all persons to public records promotes the transparency and accountability of public bodies at all levels of government. It is a fundamental obligation of government to operate openly and provide public records as expediently and efficiently as possible in compliance with this Act.

Others have focused less on the philosophical aspects and more on the practical value of transparency.

Limits on Data Sharing: Privacy, Confidentiality and Public Safety

Government data sharing has been limited by many barriers, including the costs of publishing data, the reluctance to expose poor data quality, potential liability associated with government actions, or a sense of data ownership that is not consistent with general dissemination. While many of these barriers are quite challenging, they are insufficient to justify the withholding of government data. But other barriers are fully sufficient to justify, or even necessitate, the withholding of data:

- Privacy (release of personal information such as medical records
- Confidentiality (release of trade secrets or information related to pending corporate mergers, and similar situations)
- Public Safety (release of information that could create or increase a risk to public health and safety)

Some government data must always be withheld for these reasons. But government entities at all levels should think carefully about which data sets meet these criteria, and about how the barriers might be overcome. For example, data fields that jeopardize privacy can be masked or removed from data sets about to be released. Far too often, government data has been withheld in the name of guarding privacy, even in cases where the offending elements could easily be removed.

The Practice of Data Sharing in an Evolving Environment

The Freedom of Information Act was designed to enable citizens to request physical documents from government entities. While physical documents are still in use and constitute an important information resource, the focus of data sharing has shifted to digital formats. These formats include email, text documents, databases and geospatial files. More importantly, the practice of data sharing has shifted away from stored resources such as text documents, in favor of online data. And increasingly, online data means machine-readable web service formats that can be consumed by remote web servers.

In 2000, the city of Baltimore implemented a groundbreaking data transparency program called CitiStat, which has inspired many similar initiatives across the world. CitiStat opened the city's operational databases for public access via RSS feeds and online searches (http://www.baltimorecity.gov/Government/AgenciesDepartments/CitiStat.aspx).

One of the most ambitious successors to CitiStat was the District of Columbia's CapStat, which included data from several counties surrounding the District (http://capstat.oca.dc.gov/).

In 2003, Utah's Chief Information Officer, Phil Windley, published his Web Services manifesto[2], a seminal document that directed Utah state agencies to begin making their data available as machine-readable XML web services. Web services are important because they allow web servers to request data remotely, and display that data in any number of web applications hosted by many entities. Web services allow data from multiple sources to be "mashed up" into a single application.

The transition from the sharing of physical documents, to digital documents, to online resources, to web services has greatly increased the value of government data. Today, web programmers can create almost infinite

varieties of web applications based on data sourced directly from the most authoritative providers. The flood of app contests hosted by municipal and state governments provides clear evidence of the tremendous range of possible web resources based on public data.

Data Sharing and Transparency in the Chicago Region

By 2003 the Chicago region enjoyed free or low-cost access to important public data sets, including address-level records from the Cook County Recorder of Deeds, business licenses from the City of Chicago, and state-issued licenses from the Illinois Department of Financial and Professional Regulation.

In 2005 CMAP played a leading role in a MacArthur Foundation-funded initiative called IDEA (Illinois Data Exchange Affiliates). IDEA included the City of Chicago, the Cook County Assessor's office, Metro Chicago Information Center, Center for Neighborhood Technology and Metropolis 2020. IDEA's core principles were as follows:

- Society works best when information is generally available
- Government works best when information is shared across divisions
- Web technology gives unprecedented opportunities for making data available
- Ensuring access to public data requires clear guidelines on how, when and with whom data is to be shared

IDEA promoted data sharing as a means to better government, citing these benefits:

- *Efficiency*: Information is an under-utilized asset in government. If it's not shared, the organization is not benefiting from its full value.
- *Better service delivery*: Web services can enhance delivery of government services. Residents are now challenging government, looking for convenience, looking for online access to government services.
- *Leveraging citizen participation:* releasing data helps to leverage an entire set of developers, product managers and other people interested in building applications with that data
- *Breaking down silos:* Silo-smashing is a fundamental benefit that web services can bring to government, because they allow integration of departmental data systems.
- *Better decision-making:* If decision-makers have good information in front of them, they can make better decisions about where to direct resources.
- *Government transparency:* accountability to the public is heightened when government data is opened for public scrutiny and public use.
- *Emergency response:* Lack of interoperability has been frequently cited as a contributing factor that hampered government responses to the 9/11 and Katrina disasters. Web services can deliver interoperability. No other current technology has a realistic chance of linking many disparate units of government in real time.
- *Global competitiveness:* Information is a rich asset for business and government. Information can be used within the region to foster innovation and raise the quality of goods and services. It can strengthen the capacity of our region's businesses and other institutions to compete successfully in the global marketplace for customers, capital and people.

(from http://data.cmap.illinois.gov/chidataexchange.net/Index.htm)

Best Practices for Local Governments

The following recommendations are intended to guide the data transparency efforts of local units of government—county, municipal, township and others—in the Chicago region. The items are listed in no particular order.

(1) **Release everything.** Start from the assumption that all data generated by your agency should be published online and updated regularly. If something is important enough to be tracked by your staff, it's probably important enough to be posted on the web. This reflects the "open by default" principle advocated by former federal CIO Vivek Kundra. Of course, not all data sets have equal value, so start with high-value ones (more on that below). And it's okay to post the data in summary format—although in many cases the raw, granular data has greater value to the public.

(2) **Protect privacy and public safety—and then post the data.** When data sets contain information that could jeopardize privacy or public safety, consider suppressing the offending fields and releasing the cleansed data set.

(3) **Strive for quality.** Follow the 8 principles of open government data listed at http://www.opengovdata.org/home/8principles.

(4) **Publish an inventory of all data sets** maintained by your organization, even those that have not yet been made publicly accessible. This will give your constituents a clear idea of what activities your divisions are active in, and whet information resources are generated as a result.

(5) **All FOIA requests deserve a timely response.** Either promptly comply, or promptly cite specific reasons why the request was rejected. As a best practice, follow the higher FOIA standards of states like Illinois, where government agencies bear the burden of demonstrating that a request is unreasonable. In many states, the burden is on the requester to demonstrate a compelling need for the information.

(6) **Transparency costs money,** so log staff time and other costs of providing data and information to the public and to other government entities. Log all FOIA requests and the time spent on each one. Document software costs, hours spent on improved data entry, and similar expenses. Take every opportunity to raise awareness of these costs—and the benefits of doing so.

(7) **Start small, but start now** if you haven't been posting data online. If your agency does not have a website that can serve as a data publication site, create something simple using Wordpress, Drupal or even a free blogging site like blogger.com.

(8) **Don't get locked in** to software systems that don't allow you to export your data on a regular basis. It's your data, and you need it to be available for many different applications. When you evaluate software systems for purchase, a system that doesn't allow regular exports to standard file formats should be immediately disqualified.

(9) **And don't get locked in** to software contracts that span many years, because the information technology market changes on a daily basis. It's fine to pay annual renewal fees for products that still serve their purpose as well as competing solutions. But that's different than contractually locking in for more than 3 years.

(10) **Use a competitive bid process** when purchasing software or hiring consultants. It may be easier to buy a product that you are familiar with or hire someone you know to handle your systems. But the information technology marketplace is constantly evolving and highly competitive. A bid process takes time, but can open your eyes to possibilities you hadn't considered.

(11) **Reach out to web developers** who can turn your data into something more usable and interesting. Learn what data they value most, and ask about the best delivery mechanisms. You cannot possibly provide all the data applications that people would like to have. Let the marketplace of ideas select the most useful apps, and let developers deliver those apps. Only you can provide the raw material—data—but many others can turn the raw material into useful products.

(12) **Give away your most valuable data assets.** The more you invest in a data resource, the more value it is likely to have for the public. Many government entities treat their costliest data resources as potential revenue streams, but this is seldom advisable. It's difficult to police the usage and dissemination of digital assets, and the number of paying customers tends to be small. On the other hand, the value to society generated by freely available data can be very great, if difficult to quantify.

(13) **Frequently-updated data sets deserve their own web services.** Data that changes every day or every week should not be posted as data sets, because your data consumers cannot keep their own copies up to date. This applies especially to GIS layers that are updated frequently. *NOTE:* this rule does not apply to lists that grow every day, without altering the base data. For example, a list of building permits or criminal arrests can be continuously updated without rendering the previous day's data obsolete. But updating a GIS layer to reflect changes on the ground renders all previous versions obsolete—so the change should be reflected in a web service that users can consume as needed, rather than a data set to be downloaded over and over again.

(14) **Developers love APIs.** Managing an API (application programming interface) is not simple, so it might be approached differently by large jurisdictions than smaller ones. See below for more details.

(15) **Engage the public in a transparency dialog.** Use social media and public meetings to hear from your constituents and let them hear your perspectives on the trade-offs involved when releasing information. Find out which data sets people truly value, and why.

(16) **Metadata matters.** If you release data but fail to include metadata, responsible consumers will ignore the data, while irresponsible ones will make incorrect or misleading representations of the data.

(17)**Cite the source.** When publishing data that did not originate with your agency, always cite the source, provide metadata and link to the authoritative source for the data.

(18)**Raw data rules.** As a general rule, raw data is more valuable than aggregate data, but both forms are useful. If you can release only one form, release the raw data so that consumers can generate their own aggregations.

(19)**Geocode everything.** Most government data is associated with a place on the map, and will be much more valuable if it contains sufficient spatial information to display it as a map. Of course there are exceptions. For example, the purpose of releasing employee salaries is to increase accountability to the public, not to display a map showing where government employees live.

(20) **Be a good data neighbor.** Data is most useful when it can be compared across places, so reach out to your neighbors and find ways to make your data compatible with theirs. For example, if their building permits data includes the estimated amount of work done, but yours does not, consider adopting their standard. Regional collaboration depends on shared cross-jurisdictional data practices.

(21) **Data is most useful when it can be tracked over time.** If you change the way your data is tracked, consider going backwards to make archival data comparable to your current data. Trend analyses can be powerful tools for planning, problem solving and staying competitive in a rapidly changing environment.

(22)**Embrace standards—**but at the infrastructure level, not the application level. It's crucially important that we all agree on transport-level protocols such as TCP/IP and HTTP. It's also important that we agree on a common data protocols such as XML and JSON for web services, and even schemas for common data elements. But it's seldom advisable to standardize on specific applications for tracking or displaying information. No application meets all needs, and not all government entities have the same level of resources or expertise.

(23) **Embrace open source and cloud solutions** for common data management and publication tasks. Open source tools and cloud-based software-as-a-service (SaaS) offerings can greatly reduce the cost of sharing data, and often are quite simple to use. Google Maps is a prominent example of a simple and cost-effective SaaS.

(24) **Don't reinvent the wheel-- unless you really want to.** Many common functions of local government, such as handling resident service requests, can be tracked using open standards like Open311. These open standards may not serve your needs (in which case, by all means build your own wheel) but they can provide cost-effective solutions for many common tasks.

FOR LARGER JURISDICTIONS (above 1 million population)

(1) **Maintain a data sharing portal** to handle the high volume and wide variety of data requests your constituents expect. Examples of data sharing portals include the City of Chicago's data portal at http://data.cityofchicago.org/ and Cook County's portal at http://data.cookcountyil.gov/. Data portal solutions include software-as-a-service and open-source options. Data portals typically include APIs (application programming interfaces) for retrieval of machine-readable data.

(2) **Offer space on your data sharing hub to smaller neighbors.** Smaller partners might pay their fair share of the total cost, or might simply be allowed a certain amount of space.

(3) **Provide APIs** for specialized high-value data sets. Data portals offer generic APIs for retrieving machine-readable data, but some data sets require a custom treatment. An example is the Chicago Transit Authority's API: http://www.transitchicago.com/developers/

(4) **Get your own geeks.** The debate about outsourcing vs. staffing up will continue to simmer in government circles, but large jurisdictions almost certainly can benefit from having a high-capacity information technology staff for core I.T. and systems development functions. If nothing else, you can't effectively evaluate contractors unless your people are themselves competent technicians.

FOR SMALLER JURISDICTIONS

(1) **Borrow or lease space on a data sharing hub** managed by your larger neighbors.

(2) **Join with your neighbors to increase your collective buying power.** Hardware, software and consulting services can usually be procured more cost-effectively by a consortium than a single customer. This role can be played by a GIS consortium (see below) but other kinds of consortia can convey the same advantages.

(3) **Consider joining a GIS consortium.** A GIS consortium procures GIS staffing so that all members have access to technical expertise without having to all hire their own GIS specialists. Models for a successful GIS consortium range from for-profit providers to councils of government or other local coalitions.

NOTES

1. Pennsylvania Senator William Maclay wrote in his journal: "I am now more fully convinced than ever before of the propriety of opening our doors. I am confident some gentlemen would have been ashamed to have seen their speeches of this day reflected in the newspapers of to-morrow."
2. Windley's manifesto included the following guidelines:
   1. Every data element and collection is a resource
   2. Every resource should have a URI
   3. Cool URI's don't change
   4. Preserve the structure of data until the last possible moment (i.e. return XML)
   5. Make XML Schemas available online for your XML
   6. Data queries on existing resources should be done with a GET
   7. Use POST to create new resources
   8. Document your service API using WSDL, WRDL, or some other standard
   9. Advertise the presence of the data using WSIL
   10. Adhere to data standards such as RSS where available
   11. Use Metadata (RDF) for XML
   12. Use HTTP authentication as much as possible
   13. Make data available in multiple flavors

   (from http://www.windley.com/docs/EnablingWebServices/Web%20Services%20-%20Jan%202003.pdf)