A Call to Action for State Government:
Guidance for Opening the Doors to State Data

Transparency initiatives and websites are proliferating across government and industry globally. One aspect of the transparency trend is broader access to government data. The federal government is actively developing its data portal—www.data.gov—with robust objectives regarding the number of available datasets. The full universe of outcomes is not fully understood, but it is certainly anticipated that the availability of government datasets will lead to more citizen engagement, the potential to spawn new economic opportunities, efficiencies in the creation and use of data, and greater accountability on the part of government.

These are unprecedented times regarding openness in government. President Obama has a made a commitment to open government that is unparalleled. One demonstration of this commitment is the launching of the Open Government Initiative. This initiative is founded on three principles:[1]

- Transparency – to enable greater accountability, efficiency, and economic opportunity by making government data and operations more open.
- Participation – to create early and effective opportunities to drive greater and more diverse expertise into government decision making.
- Collaboration – to generate new ideas for solving problems by fostering cooperation across government departments, across levels of government, and with the public.

"My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government."[2] *President Barack Obama*
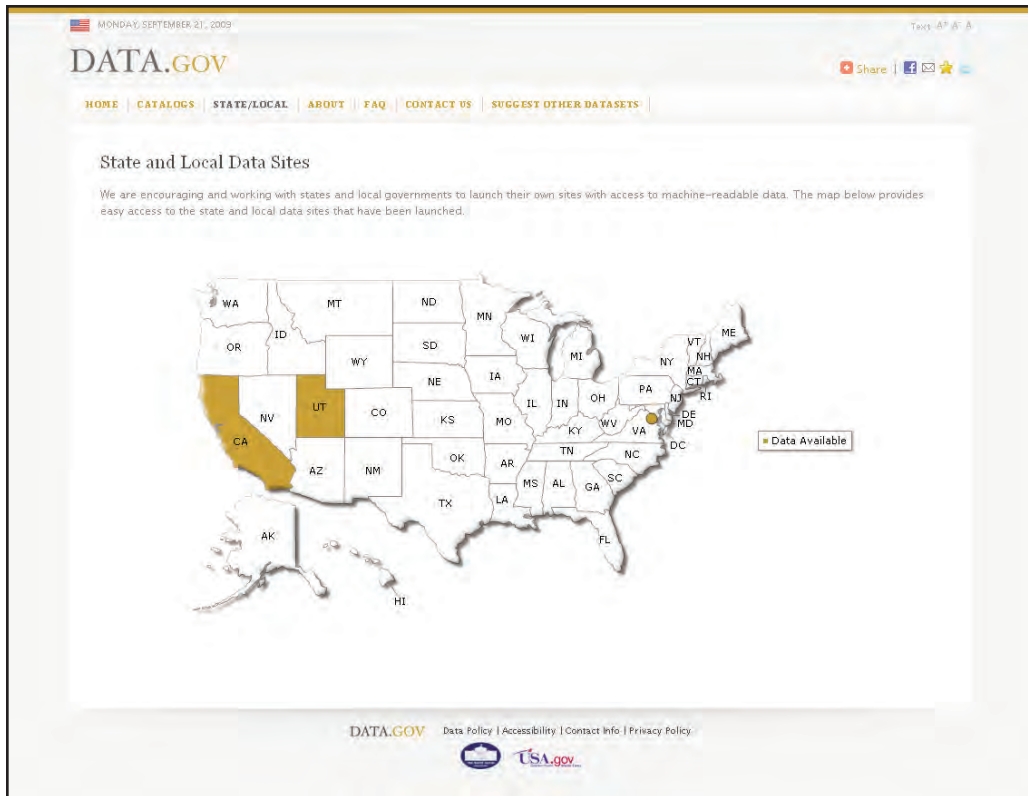
NASCIO Staff Contact:
Eric Sweden
NASCIO Senior Enterprise Architect
esweden@AMRms.com

NASCIO represents state chief information officers and information technology executives and managers from state governments across the United States. For more information visit www.nascio.org.

www.data.gov/statedatasites

State governments have been leaders in the establishment of state transparency and accountability portals and are now considering including high priority datasets in the mix.[3] The next phase in this open government transformation is the democratization of data as evidenced by the federal government's creation of www.data.gov. With vast data resources, States are in a similar position to create a new kind of relationship with their citizens. NASCIO encourages states to follow suit in the steps of the District of Columbia, the State of Utah, the State of California and several local governments.[4] The difference now is the availability and provision of raw, machine-readable data, structured to allow manipulation and analysis electronically. This allows the combining of multiple public datasets using mashups to create new information, data and services.[5] Data within the context of these discussions is unabridged data; data that has not been aggregated, summarized or interpreted.

Democratization of data doesn't stop at the states. Local governments are also pursuing open government initiatives and creating data.gov sites. The District of Columbia is the forerunner; however, San Francisco and New York City are two other models.[6] With the launching of DataSF.org, San Francisco Mayor

Newsom is hopeful that the availability of this data will lead to new economic opportunities.

Collaborate for Success

Successful planning and execution of such an initiative will require the collaboration of multiple state agencies and programs that serve as custodians or stewards of state data. These include state CIOs, chief enterprise architects, archivists, records managers, data architects, agency data stewards, librarians, portal developers, and the program management office. These roles bring with them an enterprise view relative to state data and that is the perspective that must be maintained in launching a data.gov initiative.

"Whatever else, the raw data should be made available as soon as possible. Preferably, it should be put up as Linked Data." *Sir Tim Berners-Lee*[7]

The Value of Government Data

The full list of benefits and risks associated with making state government data available in easy to find and usable formats will become more evident in the future as government and the public gain experience with this resource. State government has always been a source or supplier of public data, however it has often been difficult to discover and not available in a raw format to be consumed and reused. "Democratization of data" is often used to describe the process of making government data more available. This is now being accomplished through federal, state and local "data.gov" portals. The anticipated outcomes include: [8, 9, 10]

- Greater government accountability, credibility and integrity – because the public has easier access to government reports and the underlying data.

  Government should anticipate additional checks and balances being demanded by citizens as they scrutinize past and current legislation, and the "real" benefactors of that legislation.[11]

- More productive government staff—because of the potential for citizens to find the data they need without the assistance of government agency staff.

- More productive and effective government staff— because data is more accessible and available to government decision makers, analysts and researchers. The increased availability of data can lead to more analytical decision making versus strictly intuitive decision making. This can lead to more effective decision making with a proper balance of analytics and intuition.

- A more informed public—because the public has easier access to raw data. Evaluation of this data will lead to conclusions regarding all aspects of society and the physical environment. Citizens will become more fact based and "evidence" oriented.

  With the availability of data, policy makers, decision makers, and citizens may become more oriented and accustomed to data analysis. This means more emphasis on facts and less emphasis on intuition. This more informed public may become less accustomed to accepting the inferences of others. There may be more demand for citations of data sources. There may be more "checking up" on statements, generalities and trite comments made by public officials and others.

- More participation in government by citizens— because a data.gov portal increases the ability of the public to easily find, download, and use datasets that are generated and held by state government.

- Stimulation of innovation and new business ventures through secondary use and repurposing of data by private firms. Access to government datasets can lead to new services that harvest added value from this data.

- More and better ideas generated for improving government effectiveness—because the citizens can access government easily, they can evaluate circumstances and potentially generate ideas for solving problems and exploiting opportunities.

- More confidence in government—because citizens can see the value that government delivers.

- Higher detection of fraud and abuse—because there are more people looking for signs of it.

- Potentially new correlations and understanding that can lead to the identification of opportunities for government, business, academia and non-profits. This can be accomplished through mashups.

- Surfacing of previously unidentified, but emerging issues. This can result in more predictive analytics which can lead to a better prepared government and society.

- Identification of new opportunities for exploitation of data for good *and for bad*. Access to government datasets can lead to new value added services that harvest value from this data. The criminal element will also be looking at this data— to find ways to exploit and do harm.

- Validation or obsolescence of previous assumptions regarding *cause and effect*. And this can happen in any number of areas of interest.

- A more informed media and advocacy groups. The media may harvest this available data for substantiating issues, proposing solutions, challenging policy, or promoting a platform.

Once the data gates are open, it may never be possible to close them again. It should be anticipated that society will not only expect access to data and information, but will demand it from now on.

Federal Data.gov Site – www.data.gov

> "From the performance of a local school to the most recent statistics on crime we need to make sure that people have the facts they need to make informed choices and hold public services to account." *Tessa Jowell MP, Minister for the Cabinet Office, United Kingdom*[12]

Tools of the Trade

With the advent of data accessibility, it is quite conceivable that there will be a growing demand for capabilities for harvesting, analyzing, visualizing and presenting government data and information, and authoritative data services. These tools include the understanding and application of descriptive and inferential statistics, business intelligence, business analytics, geospatial visualization, and other visual analytics.[13]

> Visual Analytics is the science of analytical reasoning facilitated by interactive visual interfaces.[14]

For some time much of government data has been available, but one needed to know *where* to look for it. With the launching of data.gov, the federal government is paving the path so that it is easier to access federal government data. There is only one place one needs to look—data.gov. Further, given the robust objectives of the data.gov initiative, more and more data will be accessible through this door.[15] The state of Utah has "surfaced" multiple datasets that were already available for public consumption but are now easy to find. Utah has also enhanced the state's data.gov site with tools for analyzing and presenting data—and not only data pertaining to the state of Utah. The state data portal also presents publicly available data from other states such as unemployment data. In this way, data can be compared and evaluated.

Making government data more available is simply good business. It will build a stronger relationship between government and its constituents.

There is a growing capacity and sophistication on the part of citizens and organizations to develop web applications including mashups. This sophistication will increasingly put pressure on government accountability and reporting. Demands on government for custom reports—such as mashups of
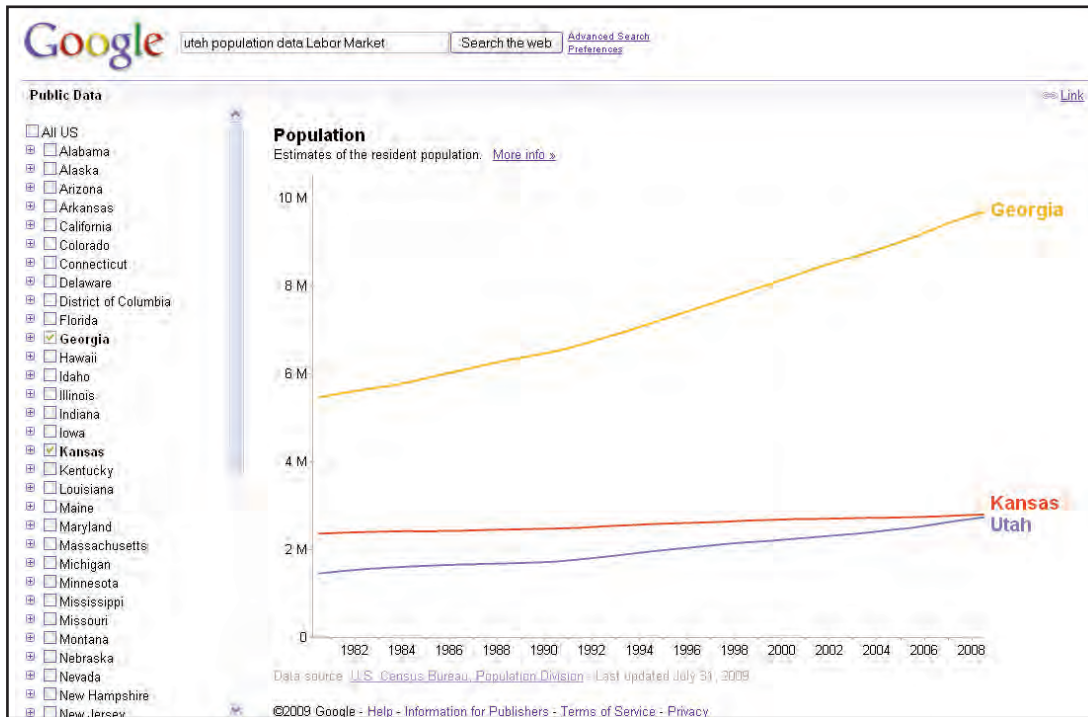
From the State of Utah data portal
(this page can be accessed from the federal data.gov site by selecting the
state of Utah under STATE/LOCAL)

bills and campaign donations – is quickly becoming impossible to meet. However, by providing the raw data to constituents, particularly through linked data, government can share this burden. The raw data can be made available, and additional value-add compilations can be achieved through mass collaboration approaches.[16] Government can't support all those specialized functions and needs, but it can facilitate their development by providing the content that makes them possible.

"It is simply unacceptable at this point in history that a citizen can use Web services to track the movies he is renting, the weather around his house, and the books he's recently purchased but cannot as easily monitor data regarding the quality of his drinking water, legislation, or regulations that will directly impact his work or personal life, what contracts are currently available to bid on for his state, or what crimes have recently occurred on his street."
Jim Willis, Director of eGovernment, the State of Rhode Island, from "Wikinomics: How Mass Collaboration Changes Everything"[17]

Principles for Open Government

In 2007, a working group was convened to establish "Principles for Open Government." The conveners included O'Reilly Media, Public.Resource.Org with the support of the Sunlight Foundation, Google, and Yahoo.[18]

This working group developed 8 Open Government Data Principles which can form the basis for state government data.gov initiatives. These principles are presented here as candidate foundational principles that should be considered for adoption and enhancement by state government.

1. **Complete**
   All public data are made available. Public data are data that are not subject to valid privacy, security or privilege limitations.

2. **Primary**
   Data are collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.

3. **Timely**
   Data are made available as quickly as necessary to preserve the value of the data.

4. **Accessible**
   Data are available to the widest range of users for the widest range of purposes.

5. **Machine processable**
   Data are reasonably structured to allow automated processing.

6. **Non-discriminatory**
   Data are available to anyone, with no requirement of registration.

7. **Non-proprietary**
   Data are available in a format over which no entity has exclusive control.

8. **License-free**
   Data are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

Additional Examples of Principles to be Considered for Democratization of Data

The following are additional candidate principles that state government should evaluate and enhance as an early step in its development of a data.gov portal. These principles are grouped into major categories.

REGULATION

- State data is a state asset and as such must be made readily available within the restrictions of applicable federal and state regulation and statutory requirements.[19]

CIVIC ENGAGEMENT

- State datasets must be treated as state government assets and managed appropriately to protect national security. As state government moves beyond currently available datasets, more scrutiny should be applied to evaluate the appropriateness of publicizing state government datasets.

- Secondary use of state government data should not include controls over its use.[20]

- Government data will not be used for harm to the state government, the public or any other entity.

- State government data availability will encourage and enhance public participation in the processes of government.

- State government data should be made available real time – or as close to real time as possible.[21, 22]

QUALITY AND SECURITY

- State government data must be protected from unauthorized modification.

- State government data must be up to date, accurate, credible, reliable, appropriate, secure and complete. That is, the quality of data presented on a state data.gov portal will be assured.

- Privacy of individuals must be respected and maintained.

- Data cannot be attested to upon secondary usage. That is, once data is downloaded, it is no longer the responsibility of state government to attest to its quality.

TECHNOLOGY

- Data should be made available digitally in structured open formats that are re-useable. Proprietary formats should be avoided. [23]

- State government data made available through a state government data.gov portal will have previously been classified as publicly available data.

- Technology barriers should not exist for access to government data. There should be as little technical restrictions as possible.

- State government data should be made publicly available without impacting the systems of records that produce and maintain that data.

- State government data will be managed using a common standardized metadata model in order to assure an established minimum information about datasets. Examples of metadata include:
  - Ownership
  - Effective date
  - Period of Coverage
  - Geographic Scope
  - Source
  - Description
  - File Format
  - Preferred Citation[24]

Metadata Model

Datasets that are made available should include additional information about the dataset in order to present the context for the data. This pedigree information includes information such as:

- the identity of the content owner and steward
- assurance that the data does not have security or privacy concerns
- date released
- format
- assurance that the dataset is in compliance with applicable privacy, confidentiality, and other relevant statutory requirements
- assurance that the data is in compliance with the information quality guidelines of the content owner / steward

A comprehensive, common standard meta model will provide the following benefits as state government moves forward with a state data.gov initiative.

- Contextual basis for interpreting datasets
- Ability to attest to the quality of the data
- Ability to trace the data back to its source
- Ability to compare data from multiple sources
- Ability to evaluate datasets for secondary use
- Consistency in quality across the states

NASCIO recommends states adopt the metadata model presented in Appendix III. This model was adapted from the federal data.gov metamodel and enhanced with additional reference to the Dublin Core Initiative.[26]

What Data to Share

The democratization of data is indeed an evolutionary process. It should be understood as an *iterative process*. Early iterations of state government data.gov initiatives should present data that is already created for public consumption, and can be anticipated as interesting and useful to the public, government, nonprofits and industry. States should not put up any data for which there are statutory privacy or security concerns.

Candidate datasets can be identified through multiple means such as:

- Collaboration with state agencies
  - What data is uniquely collected by that state agency
  - What data availability would enhance the mission of the agency
  - Identify what data is in high demand
- Search analysis on the STATE NAME.gov domain
  - This may be done using a web crawler to find publicly available datasets. HTML scraping extracts HTML code and transforms it into a more automated format that is reusable such as XML or RDF.[27]
- Common sense based on known interest and usability factors
- Implement a "suggestion box" as a means for identifying candidate data

Data.gov sites will grow in volume and sophistication over time. A first launch should not be expected to be perfect. Simply make the data available and expect future iterations to continually improve. A metamodel is important for providing important context, traceability, ownership, and quality. It has to be determined what minimum of information should be considered mandatory. It is recommended that states establish a very short list of required metadata. Pursuing a "less is more" approach will help to ensure the metamodel doesn't become a barrier to continual growth of government data transparency. In some instances very little information is really necessary to provide the best benefit to data consumers. Use of a common metamodel across the states provides a great utility for finding comparative data.

Linked Data

Government datasets can be made accessible through a variety of means including the development of a separate public data warehouse as was done in the District of Columbia. Another approach is to enable access to state government datasets through *linked data techniques*.[29] Using this approach, a state data.gov site would point to existing, publicly available datasets.

Sir Tim Berners-Lee presents the following advantages of Linked Data Techniques:

Openness:
Linked Data is accessible through an unlimited variety of applications because it is expressed in open, non-proprietary formats.

Modularity:
Linked Data can be combined (mashed-up) with any other piece of Linked Data. For example, government data on health care expenditures for a given geographical area can be combined with other data about the characteristics of the population of that region in order to assess effectiveness of the government programs. No advance planning is required to integrate these data sources as long as they both use Linked Data standards.

Scalabilty:
It's easy to add more Linked Data to what's already there, even when the terms and definitions that are used change over time.

Guidance and Recommendations

1. State CIOs should be the champion and evangelist for promoting the concept of a state data.gov portal. State CIOs should partner with agency executives, records managers, librarians, archivists, data architects and others to create appropriate leadership, visioning, and governance. Transparency must be understood as a state enterprise initiative that is enabled by state agencies in partnership with the state CIO, and the Governor.

2. State government is encouraged to develop and release data in a structured, flexible and searchable manner—as quickly as feasible.

3. Start simple by providing data that is already public. Quickly adopt governance and architecture principles and standards in order to properly plan, implement and sustain the initiative.

4. Create an inventory of currently available datasets in partnership with state agencies, archivists, librarians, and records managers. Identify high priority datasets.

5. Establish basic data categories then continue to identify data sources that fit these categories. Expand the list of categories over time.

- Financial
- Employment
- Environmental
- Health
- Energy
- Economics
- Education
- Population

6. Establish ongoing service level agreement with authoritative data owners and custodians to provide consistent and ongoing data to the portal.

7. Leverage linked data approaches to present state datasets. This simplifies the process and avoids redundant data storage.

8. Leverage work that has been done relative to creating and managing other data portals, such as a state geodata portal.

9. Develop a common metadata model that will assure consistency across the states for state government datasets.
   - Established metadata will help to ensure the quality of the data
   - Employed as part of the dataset submission process
   - Drives population of data catalogs
   - Drives search engines and other means for discovery
   - Establishes minimum requirements for the "fitness for use" of datasets
   - Adopt the metadata model presented in Appendix III of this report. Determine what elements of this data model are appropriate for the various dataset types. There may be a lighter version for certain datasets.
   - Associated metadata should be made available for consumers of datasets.

10. Collaborate with other states on what works and what doesn't.

11. Plan for ongoing marketing and awareness building related to democratization of data.

12. Plan and design feedback processes, performance metrics, and measurements of demand. Engage citizens actively to learn what data and functionality they desire in a data.gov portal. Create the necessary functionality to capture and publicize success stories from citizens, non-profits, and businesses that demonstrate the value of this new available government data. Capture, publish

and promote new business ventures that were initiated and the successes they've been able to achieve; citizens and non-profit use of government data to impact government performance and delivery of services.

13. Plan and design notification processes that use RSS feeds and other approaches to notify end users that new data is available. Provide *subscribe and notify* features.

14. Establish policy and technical governance. A policy and strategy governance organization should be initiated and sustained which includes representation from state agencies as owners and stewards of data. A technical governance organization will establish technical standards for vetting and managing datasets, dynamic links, and the design and maintenance of the portal.

15. Establish a mechanism for creating and sustaining *subject area / practice area* communities of interest.

16. Maintain compliance with existing state government standards.
    - Business
    - Process
    - Data
    - Technology
    - Accessibility
    - Security and Privacy

17. Establish and maintain an enterprise perspective. Data.gov sites will provide a comprehensive view of the state enterprise. Present this data as enterprise data with a common look and feel and functionality.

18. As state data.gov sites evolve and continue to improve, anticipate the following enhancements and issues.
    - New demands for data
    - Demand for analytical tools
    - Notification
    - Challenges in data discovery
    - Data services that are dynamic, extensible, and customizable by data consumers

19. Collaborate with state archivists, state librarians, data architects, enterprise architects and state GIS specialists in developing the strategies and plans for implementing and sustaining a state data.gov site.

20. Develop methods and procedures in cooperation with state agencies for ongoing data identification.
    - Anticipate innovative ideas from agencies and the public
    - Anticipate both structured and unstructured data
        - Documents
        - Images
        - Video
        - Audio
    - Develop methods and procedures for maintaining links
        - Detection and repair

21. Identify methods and procedures for ongoing performance metrics
    - Access frequency
    - Establish data stewardship roles
        - Data stewards will be responsible for ensuring data quality

22. Determine what data formats will be supported including:
    - Comma Separated Value (CSV)
    - Extensible Markup Language (XML)
    - Keyhole Markup Language (KML)
    - Resource Description Framework (RDF)

23. Monitor the emerging formats and presentation approaches related to data and data services.

24. Investigate innovative business models, sourcing and private sector services that enable government transparency.

Next Steps

There are a whole host of questions and issues that will eventually need to be addressed. As government moves into this new level of transparency and secondary use of government data, new issues will emerge, and numerous questions and issues will eventually need to be addressed. Expect surprises. Currently identified issues include:

- Preservation and version control.
- New modes of providing information to the public may obsolete traditional methods for receiving and responding to requests for information.
- The importance of "context" for information. How much context should be provided?
- How to validate or verify the accuracy of the content in secondary sites? It will be an issue in litigation at some point.

- Dynamic access to data. Update frequency versus dynamic links through application program interfaces. How close to real time should be expected or will be demanded by the public?
- Availability of data may be the beginning of an increase in citizen engagement to a level that is difficult to anticipate. Be prepared for new issues and challenges to state government's capacity to respond. Balancing this challenge is the anticipated increase in citizens' ability for self service.
- What tools for analyzing data will be requested or even demanded by citizens?
- With the proliferation of data, what collisions will occur as more sophisticated analysis methods emerge? What new correlations and multi-variant analysis may actually pierce the privacy and security barriers?
- Early utilization of state government data may begin with a relatively limited market. Anticipate utilization will grow. What capacity issues will emerge as utilization grows?
- Availability of data will provide the basis for comparative analysis across state governments including demographics, crime statistics, environmental quality, economic development, etc. What new behaviors and disciplines will emerge related to competition, and fact based decision making in government and industry?
- As government and society mature in their use of data, anticipate new needs and requirements that will require enhancements to the metadata model, vetting process and governance.
- Anticipate future demand for dynamic data service provisioning tools for data presentation, data analysis, and metadata searching. Increasingly, data consumers will be searching for and demanding online data services versus static, point-in-time datasets. These types of capabilities are currently available on geodata.gov.
- Understand the real personal impact on citizens and the economic impact on business that has come about, or has been enabled by this new available government data resource.
- Continue to actively engage citizens, non-profits and business through collaborative means for ongoing learning, adaptation and transformation regarding state and even national government data transparency.

Conclusions

Open government and the democratization of data is here to stay. Over time the demand for data can be expected to increase. State government should embrace this trend and enable it. The state CIO is in the unique position of viewing these resources at the enterprise level and crafting an architectural strategy and implementation plan to facilitate and enable the growth in the democratization of data. State CIOs should anticipate that making currently available datasets more available through a convenient single door is an important first step. In the near future, the state CIO can act as a champion as well as a significant promoter and visionary for new capabilities for citizens to become proactive participants in the processes of government. This can lead to the identification and development of value added tools that will assist citizens, government workers and businesses in the use of government data. The ultimate outcome is better, more engaging government, citizen partnering and ownership of their government, and realization of this great vision:

> "—that this nation, under God, shall have a new birth of freedom—and that government *of the people, by the people, for the people,* shall not perish from the earth."[30] *Abraham Lincoln*

Now technology is making this vision more possible than at any time in history. Citizens will be able to engage more effectively and easily. State government is at the brink of a new era in data democratization. This is a step in the right direction. Get ready. This adventure is just starting—and the state CIO is a key enabler.

Disclaimer

NASCIO makes no endorsement, express or implied, of any products, services, or websites contained herein, nor is NASCIO responsible for the content or the activities of any linked websites. Any questions should be directed to the administrators of the specific sites to which this publication provides links. All critical information should be independently verified.

Appendix I: Resources

Federal Data.gov site: http://www.data.gov/

Federal GeoSpatial OneStop: http://gos2.geodata.gov/wps/portal/gos

State Data.gov sites:
- The State of Utah: http://www.utah.gov/data/
- The State of California: http://www.ca.gov/data

Local Government Data.gov sites:
- The District of Columbia: http://data.octo.dc.gov/
- The City of San Francisco: http://datasf.org/

International Government data.gov sites:
- New Zealand: http://blog.e.govt.nz/index.php/nzgoal/policy-principles/
- The Power of Information Task Force Report, United Kingdom: http://poit.cabinetoffice.gov.uk/poit/

American Library Association Key Principles of Government Information: http://www.ala.org/ala/aboutala/offices/wo/woissues/governmentinfo/keyprins.cfm

The Dublin Core Metadata Initiative: http://dublincore.org/

W3C eGovernment Interest Group: http://www.w3.org/2007/eGov/IG/wiki/Main_Page
- High level description of how Linked Data can enable access to government datasets: http://www.w3.org/DesignIssues/GovData.html

Linked Data - Connect Distributed Data across the Web: http://inkeddata.org/
- Linking Open Data dataset cloud – The image presented shows some of the datasets published in the Linking Open Data community project. Clicking any of the datasets will take you to its project homepage. http://richard.cyganiak.de/2007/10/lod/

Office of Science and Technology Policy: http://blog.ostp.gov/2009/06/03/open-government-initiative-discussion-phase-transparency-principles/
- 8 Principles: http://resource.org/8_principles.html

Appendix II: Contributors

Suzanne Acar, Principal Data Architect, Federal Bureau of Investigations

PK Agarwal, IT Director, The State of California

Ed Arabas, Senior Operations and Policy Analyst, State of Oregon

Tom Baden, Chief Enterprise Architect, The State of Minnesota

Richard Burgis, The State of Michigan, Michigan Center for Shared Solutions

Terri Clark, Data Center Manager, Kansas State Legislature

Adrian Gardner, Chief Information Officer, National Weather Service, NOAA

Don Heiman, Chief Information Officer, Kansas State Legislature

Nadine M. Hoffman, Data Management, The Commonwealth of Virginia

Christopher Ipsen, Chief Security Officer, The State of Nevada

Andris Ozols, Senior Analyst, The State of Michigan

Charles Robb, Issue Coordinator, NASCIO

Doug Robinson, Executive Director, NASCIO

Ellena Schoop, Information\Data Architect, The State of Minnesota

David Strigel, Program Management Office, The District of Columbia

Rob Surber, Director of Shared Services, The State of Michigan

Barbara Teague, State Archivist and Records Administrator, The Commonwealth of Kentucky

Stacey Walker, Senior Information Analyst, The State of California

Chris Walls, Senior Graphic Designer and Web Developer, NASCIO

Chris Willey, Chief Technology Officer, The District of Columbia

Robert D. Woolley, Chief Technologist & Strategic Planner, The State of Utah

The Council of State Archivists (COSA), www.statearchivists.org

The National Association of Governmental Archives and Records Managers (NAGARA), www.nagara.org

Chief Officers of State Library Agencies (COSLA), www.cosla.org

Appendix III: DataGov Metadata Template for Datasets and Tools

NASCIO Metadata Template for State Government Dataset and Tool Catalogs

Element Number	Element	Occurrence [min,max]	Description
1	Unique ID	[1,1]	For Data PMO Use Only - will be generated when placed in catalog
2	Title	[1,1]	Unique name of the dataset or Tool. (e.g., Current Population Survey, Consumer Price Index, FBI Ten Most Wanted Widget). This field will be used to populate the data catalog; the catalog will be sorted on this field. Note: if the title is not unique within the entire data.gov catalog, you will be asked to change it.
3	Dataset Group Name	[0,1]	This is an optional field This field allows agencies to provide a Group Name to multiple datasets and/or tools in order to show that they may be presented as a group or a set.
4	Description	[1,1]	Please provide a detailed description of the dataset or tool (e.g., an abstract) such that the user would be able to determine the nature and purpose of the data. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
5	Owner	[1,1]	Entity, person and/or organization that is responsible and accountable for dataset. This may be different from Originator and publisher.
6	Originator	[1,1]	The name of the organization or individual that developed the dataset.
7	Point of Contact - Role	[1,1]	Name, title, phone, and email address of where questions from the Data.gov PMO should be sent. Note that this information will not be exposed in the Data.gov interface but may receive public email comments made specifically about this dataset or tool.
8	Compliance with Agency's Information Quality Guidelines	[1,1]	Confirm that the dataset meets the Agency's (as identified in Element 6) Information Quality Guidelines (Yes, No). If the dataset is not in compliance with your Agency's Information Quality Guidelines it will not be posted on Data.gov.
9	Data.gov catalog type	[1,1]	Select the appropriate catalog on Data.gov that this submission should be listed (i.e. Raw Data Catalog or Tool Catalog).
10	Subject area (Taxonomy)	[1,1]	Please choose the category from the drop-down menu that best describes your dataset. If more than one category applies, choose the category that you think most people would use. We realize that this taxonomy, which is based on the Statistical Abstract, is not perfect for all datasets. If the item being described is a 'data mining and/or extraction tool," then please provide the category that describes the underlying dataset. FEA - The Interior's DRM (Data Reference Model) extends the FEA DRM with the department's Conceptual and Logical Entity Relationship Model Diagrams that are categorized within the DOI Subject Areas and Information Classes.
11	Associated Datasets	[0.n]	Information about other, related datasets that are likely to be of interest (e.g. derived from common sources, similar theme, subject, etc.)

Element Number	Element	Occurrence [min,max]	Description
12	Specialized data category designation	[1,1]	Identify the type of dataset (i.e., administrative, statistical, geospatial, surveillance or research). Please choose the category that best fits the dataset, understanding that there is some potential for overlap among these categories. Note that some types of data have additional metadata requirements (e.g., see statistical datasets). If the item being described is a 'data mining and/or extraction tool," then please provide the designation that describes the underlying dataset from the pre-populated dropdown list. Note: If you select Geospatial, you should submit your dataset to http://geodata.gov and you do not need to fill out any additional information for this submission. We would, however request that you submit the elements 1 through 7 to datagov.submit@gsa.gov so that we can synchronize our efforts with geodata.gov. If you select "statistical", please note the section "SM" that follows.
13	Keywords	[1,n]	Searchable keywords help users discover your datasets from different perspectives. They also provide ways of identifying other datasets that are similar to yours. Please include terms that would be used by both technical and non-technical users. If the item being described is a 'data mining and/or extraction tool," then please provide keywords that describe the underlying dataset. Agencies are encouraged to include as many keywords as possible. Please use commas to separate keywords.
14	Publication Date	[1,1]	Date when dataset was first made available to the public.
15	Date updated	[1,1]	Date of last change to dataset or tool.
16	Agency Program URL	[1,1]	URL that is closest to the program that is responsible for this dataset or tool.
17	Agency Data Series URL	[1,1]	URL to the high level repository in which this dataset or tool resides at the agency (e.g., the HTML page that displays the link to the dataset).
18	Collection mode	[1,n]	How the data is collected. Is this a product of an operational system? Was it derived from a survey? If the latter how was the survey conducted?
19	Maintenance and Update Frequency	[1,1]	The frequency with which changes and additions are made to the dataset after the initial dataset is completed.
20	Period of Coverage	[1,n]	Dates or time interval(s) covered by the data. Please use commas to separate multiple periods of Coverage.
21	Unit of analysis	[1,1]	The unit of analysis is the major entity that you are analyzing in your study (e.g., person, household, forest, county, establishment). If the item being described is a 'data mining and/or extraction tool," then please provide the category that describes the underlying dataset.
22	Spatial Extent of Data	[0,n]	Describe the geographic are covered by the dataset. Please indicate the geographical extent covered by this dataset. In the case of multiple locations, please delimit with commas. Some datasets are not earth-based, thus will not use this field (N/A). You may use commas to separate multiple entries.

Element Number	Element	Occurrence [min,max]	Description
22.2	Geographic Granularity	[0,1]	This is an optional field. Please indicate the most detailed level at which the geography is defined (e.g. City vs. zip code vs. longitude/latitude pair).
23	Data dictionary/variable list	[1,1]	URL to resource containing variable names, descriptions, standard vocabularies and taxonomies, units, multipliers, etc. May be identical to element 23.
24	Data collection instrument	[1,n]	URL for resource containing a copy of, or detailed descriptions of, the data collection instrument for each listed mode. May be identical to element 23. Multiple URLs should be separated by commas.
25	Metadata Standard Name	[1,1]	This field may be used when others make reference to the data, as in a bibliographic citation or source reference. If the agency does not have a standard reference for this dataset, simply provide the URL for the dataset . The name of the metadata standard or guidelines used to document the dataset.
26	Number of Datasets Represented by this Submission	[0,1]	This is an optional field. If this submission is a compressed file, data extraction tool or mining tool, please enter the total number of datasets represented by this submission.
27	Additional Metadata	[0,1]	This is an optional field Please provide a URL to any additional metadata for the dataset or tool.
Downloadable file specific metadata (repeat this section for multiple download formats using multiple columns in the following rows.			
D1	Access point	[1,1]	If the dataset is downloadable, enter the URL for instant access to the downloadable data file. This is the URL for access to the dataset via a "one-click download". Please enter a URL only.
D2	Media Format	[1,1]	In some cases files are downloaded in a compressed file (e.g. zip). Please enter the media type for information contained within the compressed file (RSS, XML, CSV/TXT, XLS, Shapefile, KML/KMZ) If not compressed, enter the file suffix of the downloadable file.
D3	File size	[1,1]	If downloadable, please enter the size of file in MB. Should be limited to 15 characters.
D4	File format	[1,1]	If downloadable, enter the format in which file may be downloaded. For Raw Data Catalog, select from the following options: RSS, XML, TXT (CSV), XLS, KML/KMZ, Shapefile, or map. For Tool Catalog, select either Data Extraction Tool or Widget.
Additional Statistical Metadata required for Statistical Datasets			
SM	Statistical methodology	[0,n]	Components identifying statistical information/properties of data.
SM.01	Sampling	[0,1]	Description or URL of resource containing more detailed information.
SM.02	Estimation	[0,1]	Description or URL of resource containing more detailed information.

Element Number	Element	Occurrence [min,max]	Description
SM.03	Weighting	[0,1]	Description or URL of resource containing more detailed information.
SM.04	Disclosure avoidance	[0,1]	Description or URL of resource containing more detailed information.
SM.05	Questionnaire design	[0,1]	Description or URL of resource containing more detailed information.
SM.06	Series breaks	[0,1]	Description or URL of resource containing more detailed information.
SM.07	Non-response adjustment	[0,1]	Description or URL of resource containing more detailed information.
SM.08	Seasonal adjustment	[0,1]	Description or URL of resource containing more detailed information.
SM.09	Data quality (variances, CVs, CIs, etc)	[0,1]	Description or URL of resource containing more detailed information.

Notes: This metamodel was adapted from the federal Datagov Metadata Template Release 2 v1 5 07-16-09 and enhanced with further reference to the Dublin Core Initiative.

Appendix IV: Endnotes

[1] From the White House – Open Government Initiative website, http://www.whitehouse.gov/open/about/.

[2] "Transparency and Open Government", President Barack Obama, January 21, 2009. Retrieved on August 23, 2009, from http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government/.

[3] See NASCIO's website for a catalog of state government transparency websites. See http://www.nascio.org/advocacy/recovery/.

[4] See www.data.gov. Select "STATE / LOCAL" to see data portals for state and local government.

[5] "Mashups Using Government Data", Preserving State Government Digital Information -Minnesota Historical Society, Hoffman, N., March 30, 2009. Retrieved on August 25, 2009, from http://74.125.95.132/search?q=cache:8sg3KVllmbAJ:www.mnhs.org/preserve/records/legislativerecords/docs_pdfs/DataMashups1.pdf+Mashups+Using+Government+Data&cd=1&hl=en&ct=clnk&gl=us.

[6] "Open Government Agenda Spills into States and Localities", The White House – The Blog. Posted by Robynn Sturm on August 27th, 2009 at 6:47 p.m., retrieved on August 28, 2009, from http://www.whitehouse.gov/blog/Open-Government-Agenda-Spills-into-States-and-Localities/.

[7] "'Get the raw data out there ASAP': Berners-Lee" by Tom Espiner, Public Sector News. Retrieved on August 22, 2009, from http://www.silicon.com/publicsector/0,3800010403,39445943,00.htm.

[8] "The Value Proposition of Transparency" Influence, March 5, 2009. Retrieved on August 22, 2009, from http://influence.forumone.com/archives/332-The-Value-Proposition-of-Transparency.html.

[9] From the federal data.gov website. Retrieved on August 22, 2009, from http://www.data.gov/about.

[10] "Why Transparency? 13 Risks, 27 Rewards and 5 questions". Retrieved on August 22, 2009, from http://productfour.wordpress.com/2009/03/14/why-transparency-13-risks-27-rewards-and-5-questions/.

[11] "Under the Influence: 60 Minutes' Steve Kroft Reports On Drug Lobbyists' Role in Passing Bill That Keeps Drug Prices High. 60 Minutes interview with Rep. Walter Jones (R-N.C.), left, and Rep. Dan Burton (R-Ind.). Segment was originally broadcast on April 1, 2007. It was updated on July 23, 2007. Retrieved on September 8, 2009, from http://www.cbsnews.com/stories/2007/03/29/60minutes/main2625305.shtml?tag=contentMain;contentBody.

[12] "Pioneer of the World Wide Web to advise the government on using data". Cabinet Office, United Kingdom. The Cabinet Office has an overarching purpose of Making government work better. Retrieved on August 23, 2009, from http://www.cabinetoffice.gov.uk/newsroom/news_releases/2009/090610_web.aspx.

[13] Thomas, J., Cook, K., *Illuminating the Path – The Research and Development Agenda for Visual Analytics*, IEEE Computer Society, ISBN 0-7695-2323-4, Los Alamitos, CA. Also see The National Analytics Center http://nvac.pnl.gov/.

[14] Thomas, J., Cook, K., Illuminating the Path, p. 4.

[15] "OMB reveals transparency roadmap", Federal News Radio, May 22, 2009. Retrieved from http://www.federalnewsradio.com/?sid=1681070&nid=35 on August 20, 2009.

[16] Tapscott, D., and Williams, A.D., *Wikinomics: How Mass Collaboration Changes Everything*, Penguin Books, London, 2006. p. 199-200.

[17] *Wikinomics*, page 200, interview with Jim Willis.

[18] Open Government Working Group, December 7-8, 2007, convened by O'Reilly Media and Public.Resource.Org in partnership with Sunlight Foundation, Google and Yahoo. Retrieved on August 31, 2009, from http://public.resource.org/open_government_meeting.html.

[19] "Data Governance – Managing Information As An Enterprise Asset Part I – An Introduction". NASCIO, April 2008. www.nascio.org/publications.

[20] See Federal data.gov Policy Statements. Retrieved on August 22, 2009, from http://www.data.gov/datapolicy.

[21]   "Democratizing Data: Author David Stephenson Seeking Tableau Customers to Tell Their Stories" by Elissa Fink on February 24, 2009. Retrieved on August 22, 2009, from http://www.tableausoftware.com/blog/democratizing-data-w-david-stephenson-seeking-case-studies.

[22] "Open Government Data Principles", created December 7-8, 2007, at the Open Government Working Group Meeting in Sebastopol, CA, convened by O'Reilly Media and Public.Resource.Org in partnership with Sunlight Foundation, Google and Yahoo. Retrieved on August 31, 2009, from http://resource.org/8_principles.html.

[23] Hoffman, N., "Mashups", p. 3.

[24] See Federal data.gov Policy Statements previously cited.

[25] "The Open Sweden Campaign", May 13, 2002. Retrieved on September 21, 2009 from http://info.worldbank.org/etools/library/latestversion_p.asp?objectID=7293&lprogram=1.

[26] The Dublin Core Metadata Initiative is an open organization engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include work on architecture and modeling, discussions and collaborative work in DCMI Communities and DCMI Task Groups, annual conferences and workshops, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices. See http://dublincore.org/.

[27] "The Future of Government Is No Government", Gartner, April 24, 2009. ID Number: G00167368.

[28] "Internet Overtakes Newspapers as News Outlet", Pew Research Center Publications, December 2008. Retrieved on August 23, 2009, from http://pewresearch.org/pubs/1066/internet-overtakes-newspapers-as-news-source.

[29] "Putting Government Data Online", Sir Tim Berners-Lee, W3C, makes the case for using linked data techniques. Retrieved on August 24, 2009, from http://www.w3.org/DesignIssues/GovData.html.

[30] Source: *Collected Works of Abraham Lincoln, edited by Roy P. Basler*. The text above is from the so-called "Bliss Copy," one of several versions which Lincoln wrote, and believed to be the final version. Retrieved on August 23, 2009, from http://showcase.netins.net/web/creative/lincoln/speeches/gettysburg.htm.