Testimony before the

ASSEMBLY STANDING COMMITTEE ON GOVERNMENTAL OPERATIONS
ASSEMBLY STANDING COMMITTEE ON OVERSIGHT, ANALYSIS, and
INVESTIGATION
LEGISLATIVE COMMISSION ON GOVERNMENT ADMINISTRATION

Hearing on the implications of the information revolution for state
government.

May 22, 2012

Statement of

Professor James A. Hendler

Tetherless World Professor of Computer and Cognitive Science
Rensselaer Polytechnic Institute

Trustee, Web Science Trust, UK
"Internet Web Expert" to the US Data.gov project

Summary of Recommendations

The State of New York should become a leader in Open Government Data Sharing in
several ways:

- The State of New York should move rapidly towards making State data
  available online in machine-readable formats.
- The State of New York should make guidance available to both State and
  Municipal agencies on best practices for data sharing.   These should include
  suggestions for best formats for different kinds of data, standard "naming"
  conventions, and privacy protection in the released data.
- The State of New York should provide federated cataloging capability for
  counties, municipalities, and tribes within the State to encourage further
  open government data sharing by those entities
- The State of New York should explore mechanisms by which the "crowd
  sourcing" of software development, data improvement, and data usage can
  be encouraged and supported.

Statement:

Chairman Englebright, Chairman Hevesi, Chairman Latimer, and members of the Committees on Governmental Operations and Oversight, Analysis, and Investigation; and the Commission on Government Administration, it is an honor to be asked to testify at today's hearing. My name is Jim Hendler and I am the Tetherless World Professor of Computer and Cognitive Science at Renssealer Polytechnic Institute in Troy. I also serve as a trustee of Web Science Trust located in the United Kingdom and serve as an "Internet web expert" for the US government's Office of Science and Technology Policy.

I believe the role of modern information technologies in improving governance, transparency and especially innovation in our State is an increasingly important one, and I appreciate being allowed to share my opinions with you on this matter. The purpose of my remarks today is to make recommendations that the State could take to improve its efforts in the Open Government area, and particularly in the area of the sharing of New York State Open Data.

To provide a bit of background before I start, I am a New Yorker by birth, born and raised in Queens, and I grew up in the New York City area. I spent much of my professional career working at the University of Maryland, but was extremely pleased when an opportunity to move to this region was provided to me by RPI, and since 2007 I have been a professor there.

My academic background is in computer science, but since the early 1990s I have been working on Web technologies. I am known as one of the inventors and innovators in what are now known as Semantic Web or sometimes Linked Data technologies, a technology I also helped to nurture when I spent three years as a Program Manager, and later chief scientist, at the Information Systems office of the Defense Advanced Research Projects Agency (DARPA). Since leaving DARPA in 2001, I have served on several advisory panels for the Department of Defense and NASA, and, in 2002, I was awarded an Exceptional Civilian Service Medal by the US Air Force in honor of this work. (A full curriculum vita with all of my publications, talks, awards and grants is available online.[1])

Since coming to RPI, I have become increasingly involved in working with government bodies on the release of open government data. My research group developed technologies for making the datasets being shared on the US open government web site, Data.gov[2], more accessible to both developers and end users. In 2010, the White House acknowledged this work, and I was asked to provide expertise to the Data.gov project, as well as the more general Open Government Initiative, in the role of an "Internet Web Expert" working with the Office of Science and Technology Policy and the General Services Administration. As well as working with the US government, I have been asked to provide expertise to a number of

---

[1] PDF: http://www.cs.rpi.edu/~hendler/NormalVita.pdf
[2] http://Data.gov

governments, working in various ways with the governmental partners in the UK, India, Kenya, and Brazil, and for various states and cities in the US, including informal discussions with the New York State government.

The focus of my remarks today will be on the important role that Open Government Data plays in the growing open government movement. Where many open government sites focus on making reports, documents or schedules available, open data sites make the databases collected by government agencies accessible in machine-readable forms. In areas such as health, environment, transportation and education, the role of data in providing guidance to policy makers is well known. By releasing such data, applications in these, and other important areas, can be used by third-party developers, both inside and outside the government.

Currently, governments around the world and throughout the US are making more and more of this kind of data available. In a project being conducted by my research group at RPI, we have identified over 850,000 open databases from more than a hundred fifty governments (federal, state and municipal) around the world. In the US alone we have found a combined total of over 448,000 from the Federal government, over thirty states and a number of larger cities and tribal entities.

A little later in my remarks I will discuss the New York State efforts with respect to open data, and make some recommendations, but first I'd like to spend a moment exploring why so many governments are sharing their data. In particular, I am often asked whether the cost of creating data sharing sites generates appropriate return on investment. While definitive answers to that question are a matter of current concern in academic circles, there is already a growing body of literature showing that the answer is "yes." And not just in financial ways, but also in terms of improved transparency, innovation, and citizen involvement.

In the US, in 2010, then White House special counsel Norman Eisen was quoted as stating that there had been a significant drop in Freedom of Information Act requests, and that the cost savings was considerable.[3] He attributed that savings to US data sharing efforts. That same year, the Harvard Business School released a study examining the business case for Data.gov[4] and this study was amended in a 2012 document by Vivek Kundra, the former US Chief Information Officer who was working as a fellow at Harvard.[5] These papers conclude that there is:

- Return on investment from innovators who develop third party applications using the government's data,
- Increased transparency, particularly in the area of "Data Journalism" which can turn the data into visualizations for increased citizen understanding of government,

[3] *cf.* http://voices.washingtonpost.com/federal-eye/2010/06/post_12.html

[4] PDF: http://**data.gov**/documents/**hbs_datagov**_case_study.pdf

[5] http://www.slideshare.net/govloop/digital-fuel-of-the-21st-century-innovation-through-open-data-and-the-network-effect

- Cost savings within government as funding decisions are increasingly backed by open government data obviating the potentially biased claims of special interests, and
- Increased opportunities for direct citizen involvement in the public sphere.

In the United Kingdom, which many perceive as taking a lead in the quality of their open government data sharing, in part due to US funding cuts to our efforts, there have been similar studies. One ongoing effort at Southampton University, for example, is documenting the comparative return on investment for local councils and the federal government. Early results appear to show that the more local efforts have even greater immediate returns than do the federal efforts.

This latter result, although still primarily anecdotal, argues that in the US efforts on the part of States, counties and municipalities have great potential. This is borne out by the returns that have been seen on the part of many of our nation's larger cities, especially New York City which is viewed by many as the world's best municipal data-sharing site. NYC agencies have claimed increased innovation on the part of developers – this was spurred on in a conscious way by contests and "hackathons" which brought developers and municipal employees together to explore possible applications. This resulted in those civilian developers creating applications that would previously have required time-consuming and expensive municipal acquisitions. It also led to better citizen engagement, positive press response, and many other benefits. The response in city government has been very positive, resulting, in February of this year, in the passage of the city's Open Data Law, which had four main objectives:

- Empowering the Department of Information Technology and Telecommunications (DOITT) to develop standards and processes for bringing all of NYC's quantitative data to a single Web portal,
- Creating an inventory of all datasets,
- Setting a timeline for agencies to put the quantitative data online, and
- Creating a feedback process by which the public can express concerns and interests with respect to the datasets and what they show in terms of municipal performance.

New York City is clearly a special case, and what happens there is not necessarily reproducible to the other municipalities in New York or to the State itself, but it is clear from the level of effort, and this legislation, that they have found great benefit in the sharing of data.

We now come to the crux of my remarks: what is the status of data sharing by the State of New York? Unfortunately, despite our size and prominence as a state, we are not a leader in this area. While the overall open government activities of this legislature are quite strong, data-sharing lags behind. The New York State Senate, for example, under a previous CIO created a then state-of-the-art website for

sharing data in 2009.[6] Since then, however, the site has been maintained, but not significantly improved. The Empire State Development Agency maintains a "data center" that does contain interesting data on the State, but much of it is collected from Federal sources (for example Census data), rather than reflecting the data collected by State agencies. Additionally, a number of private, non-governmental entities publish state data, but generally without access to the public data that the state has collected, or with only limited access thereto. While it is hard to rate State data-sharing sites, it is clear that New York lags behind many of the thirty-three states that share their data, and also behind that of its largest city.

What can the State do about this? An obvious recommendation is for the State to increase its efforts to catch up with, and eventually lead, the other states participating in Open Government Data initiatives. As should be clear by this point, the benefits to government agencies, developers and innovators, and public/private partnerships provide powerful incentives to pursue this course.

But how can this be achieved? I believe there are some important steps that the State should pursue, many of which are easy to do and can be started without great investment. There are several that I believe would have immediate benefits.

As the ultimate goal is to have a central site where data can be easily found and reused, the starting place is making the data available in forms which modern information technologies can take advantage of. To this end, the State of New York should move rapidly towards making State data available by encouraging our agencies to put their data online in machine-readable formats. Policy and legal guidelines for doing this are starting to be available, and New York State could base its efforts on those.   For example, in the Federal Government, the Digital Accountability and Transparency Act of 2012 (H.R. 4126) is now moving through the Congress. The New York City Open Data legislation is, of course, another example that could be used.

While the above is easy for a technologist like me to say, it is clearly a challenge in many cases for agencies to know how to go about doing this. I would suggest that the State of New York should produce, and make available to State (and other) agencies, guidance on best practices for data sharing.   These should include suggestions for best formats for different kinds of data, standard "naming" conventions, and privacy protection rules for the released data. The State could easily develop these guidelines by working with a combination of those governmental and private entities, including academia and industry, which are already leading the efforts I alluded to earlier.

To further encourage data sharing not just at the State level, New York should make sure that its data sharing efforts provide the capability for counties, municipalities, and tribes within the State to participate in the overall open government data

---

[6] A description of the development of the site is available at
http://globehoppin.com/2010/10/13/open-senate-overview/

sharing. By supporting a site that can provide means for "federating" the data collections of agencies throughout the various levels of government in the State, data journalists, developers and concerned citizens could find that data which most effects them. Just as one example of the benefits of this, companies interested in relocating would be able to learn more economically about sites throughout New York, increasing our attractiveness as a destination.

Finally, in keeping these remarks short, I have focused primarily on the benefits of the release of public data, rather than focusing on another aspect of open government data, which is the benefits to data collection and curating that can result from the "crowd sourcing" of data. In her book "Wiki Government," Beth Noveck, former Deputy CTO of the United States and a New York resident, discusses this at length.[7] The approach she advocates has led to new applications in improving health, creating new business opportunities, reducing crime, and improving environmental conditions around the world. New York State should explore mechanisms by which this crowd sourcing approach can be encouraged and supported for the benefit of our citizens.

In summary, I think the capabilities exist in New York to develop and maintain a leading presence in government data sharing. The benefits to the State would be seen in a number of ways that I have discussed, and the development surrounding this release would have numerous positive impacts on the State and its residents.

I thank you again for this opportunity to testify and look forward to following up with any questions you may have now or in the future.

---

[7] Beth Noveck, *Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful,* Brookings Institution Press:DC, 2009.