

A Digital Government Research Agenda for Future Partnerships
Between
Statistical Agencies , Academia and NSF

Charles J. Rothwell , Associate Director, National Center for Health Statistics

Introduction:

The following comments derive directly from a work group I facilitated at a NSF Digital Workshop this April. The workgroup consisted of representatives of Federal statistical agencies and PI' s of Digital Government projects, which have statistical agencies as partners. The work group was charged with proposing some next steps in our research partnerships. Thus all insightful comments I make derive from that work group; however, I do take responsibility for any muddled headed ideas.

The initial IT research partnership between academia and Federal statistical agencies supported and nurtured by the Digital Government Program of the National Science Foundation can be considered a success on many levels. Funded research in new data collection technologies; new techniques to present tabular and graphical information to the public; technologies to ease access and virtually bring together the data holdings across federal agencies; technologies which allow greater access to data while providing improved security and confidentiality; new geo-spatial display systems are all areas of immense importance to statistical agencies. Academia has found that work with Federal

statistical agencies provides fertile research opportunities in areas which can make a difference to the emerging electronic fabric which will bind our society together. Through these initial research efforts, both parties have learned how to work together more effectively and have shown that NSF can and should play a central role in supporting applied research to help government transform itself in the support of a digital society.

A Research Agenda for the Future:

A new research agenda should be built on the success of our initial work together. We should encourage additional research in the areas currently funded; use the outcomes of initial projects to help focus on related research opportunities and build an academic infrastructure which looks to the issues of governance in a digital society as a test bed for research. In that light and to help frame a research agenda for the future, I offer the following:

Research about Users: What information does the public want and how do they want it? There is considerable controversy on current methods of data collection on users of government information holdings. Examples include the recent Executive and Congressional concern over the use of “cookies” on government web sites and the public’s long standing concern over government collecting personal information. Yet government needs to be responsive to the public need for information from an access, content, display and organization perspective. The answer is not just more information, but the right information, in the right format, at the right time, to the right device(s) in an understandable manner. We need “metadata” on the users of government information. How

can we become more informed about users in a changing information society and still protect their anonymity? Can intervening technologies be developed and provided between the user and government to insure anonymity and still provide government information providers with data to improve their information services? How can we learn from users entrance vocabularies on how to structure and provide information? How can cognitive labs, focus groups and other similar settings provide a more in-depth understanding of what information various segments of our society (young, old, novice, expert, those of differing ethnic and educational backgrounds) want and how they access and process information? How do various segments of society operate their computer; what tools do they use; and on what type of devices would they like to receive this information? Do users really want portals or better search and linking engines? Cognitive methods are needed to identify users approach to tasks in the pursuit of information discovery. What kinds of collaborative tools could or should government provide to encourage electronic communities to form and utilize our information? These questions and those which follow provide a fertile ground for research and innovation.

Research about Information on Data: From discussions on barriers to the current Digital Government research activities in data access, one significant barrier identified is a lack of metadata. Metadata standards are not the only answer, especially for legacy systems. Are there a core set of absolutes within the various metadata standards which can be applied to data holdings from legacy systems? Can technologies be developed to move existing electronic

documentation into metadata holdings? Do we really know what is essential metadata to the novice user or to the expert? Technologies are needed to help the novice user utilize metadata to better structure their information request, direct the request to appropriate information from vast government information holdings and to understand what they have received.

Research on Technologies to Collect Data from Respondents and Providers: Government and specifically Federal statistical agencies sponsor the most extensive and complex surveys of citizens and institutions in the U.S. and these surveys form the basis for our information holdings which are returned to the public. For the most part these surveys are voluntary and time consuming for the respondent and yet to provide solid estimates of rare events or with enough demographic specificity, there needs to be exceptionally high response rates. Also these surveys need to be flexible in their administration and content. Technologies are needed to speed the authoring of complex surveys, their maintenance and modification; improve the interfaces of the software with the author, survey administrator and respondent; develop new modalities of data collection; and enhance the respondents understanding of the importance of the information they are providing. Examples might include: the use of voice, picture and sensor input; web-based delivery ; multi-media interfaces; easy to use survey authoring systems; cognitive studies on software interfaces for authors, administrators and respondents; etc.

Research on Technologies to Improve Data Access and Protect

Confidentiality of Respondents: Many statistical agencies provide micro data files derived from their surveys to researchers, policy makers and the public. Because of the complex survey designs, the development of point estimates and variances are not straight forward. The ability to undertake queries on these files is therefore limited and little capability is now available on the web to access multiple data files across many agencies. Technologies to wrap around existing micro-data files for fast inquiry and graphical display utilizing appropriate analytic techniques for the novice user are needed. A variety of new visualization techniques need to be developed to help in data mining. For statistical organizations which can not release micro-data files or for those who can only provide partial releases, technologies need to be developed which will allow for remote access to these files for analysis without jeopardizing the confidentiality of the respondents.

Research on Technologies to Help Improve Information

Competencies of an Evolving Information Society: The information derived from Federal surveys and administrative systems are complex and for the novice or occasional user are difficult to understand, compare and combine into a useable packet of information. We need to find better ways to depict through the use of intervening technologies, the concepts of uncertainty, error and data quality and allow the user to select their competency level which would in turn trigger the extent of these intervening technologies. These technologies could

also be used as learning tools to improve the information competency of the user in an unobtrusive manner.

Research on the Virtual Information Organization: Is a virtual statistical organization just a portal? Is it a portal with a few unique applications? Is a virtual agency more than just the sum of its organizational parts? Are there new combined data and information products which should be developed and provided? What are the research possibilities for technologies to access multiple data sets for geographic or demographic analysis and display? Should the virtual agency support a statistical library and collaboration center by recording and sharing of the results from queries of other users? How should government information agencies evolve to support a digital society?

I hope the list of questions and suggestions I have presented to you provides a starting point for future partnerships. We have some exciting times ahead of us in trying to find both the technology tools and the organizational incentives to transform the way government currently works.