

## GROUP FIVE:

### Digital Government Technologies: What's Next?

**Participants:** Leana Golubchik (chair), Jamie Callan (scribe), Vijay Atlira, Jamie Callan, Catherine Dippo, Ali Farahani, Leana Golubchik, Kincho Law, Scott Midkiff, Charlie Rothwell

#### DISCUSSION

The digital government program currently funds the development of technologies such as data integration, heterogeneous database access, diverse data types, security and confidentiality both in data gathering and dissemination, etc. Yet at the same time, there are a range of related technologies that could be included. This group considered which technologies need to be further developed, better exploited and, which research domains related to these technologies should be included.

#### Technology Transfer.

Technology transfer activities related to Digital Government projects are badly underfunded. Current funding is for feasibility studies and prototypes only; some government partners understand this, some do not. Graduate students are the wrong resource for building application systems. Agencies should fund technology transfers, but may not have the budget or appropriate technical or other resources. Agencies would also rather buy than develop. Current research funding does not go far enough to demonstrate commercial viability so that companies would pick up research, or so that VCs or others would fund startups. Need something like SBIRs more tightly integrated with Digital Government program. The Digital Government program needs to get a slice of the \$200 million allocated for cross-agency projects; Digital Government projects have already been peer-reviewed and evaluated in operational environments, so some sort of "fast track" status is appropriate. Projects need greater visibility beyond current DigiGov partners, to entice other agencies, or so that multiple agencies could share development costs.

#### Business Models (Business Process Re-engineering).

Even when the technology path is clear, potential new business models (i.e., new ways of doing business) for government agencies are unclear. Who is the "customer" for a system that monitors people on parole? It is the agencies that need more help developing business models they are completely overlooked, and underfunded. One example of this is: putting probation officers in schools (because that's where many of the offenders are)—is this effective? Or is it a useful or counterproductive resource for schools, how does it interact with other programs? The group feels that there should be an emphasis on pilot programs that would be widely applicable and that could be copied easily if successful (e.g., at county level). We compare this to monitoring non-violent drug offenders on probation; monitoring via postcards is ineffective (low response rate); new program is kiosk related, so participation is more easily enforced; could this be ported to the Internet. Does it make sense? Digital divide issues (does this community have appropriate access). What would be better business models for this population? One idea is that we need more Business School types to study reengineering of government services.

#### More cross-jurisdictional portals.

There is a need for more topic-specific state-wide information portals, e.g., transportation. This is partly a technology issue, partly a matter of integrating a breadth of resources. User-specific focus, as opposed to agency-specific. Might tie into business process reengineering projects. There are a lot of these kinds of projects that need to be done

#### Privacy.

We felt that if there were unique identifiers that it would be very useful. There are however, major privacy issues, such as authentication being a key issue; anonymity; many policy issues, not really the domain of CS folks; the main issue is social costs; the more interrelated systems become, the more serious are the consequences of hackers, identity-theft, etc; there is a need for people to study the ripple effects of errors, hacks, etc, in inter-related systems, i.e., how errors propagate, and how errors can be corrected

#### Large-scale integration of heterogeneous resources.

Current projects rely heavily on unified schemas, global ontologies, and manual wrapping of databases. We feel that

these projects will become the Digital Government success stories because the advantage would be to have a one-stop shopping and integration of services. The disadvantage of course is that these approaches don't scale (economically) to large numbers of agencies, because manually wrapping hundreds or thousands of resources is not practical (expensive). Also, integrated systems are even more difficult to change in the future, because a local change ripples into other systems. The success stories will be difficult to reproduce on a large scale. We need more emphasis on automatic wrapping of resources and automatic use of ontologies, as USC/ISI project is doing; this is a hard problem, but really the only approach that is viable in the long-term and on a large scale.

#### **Scalable infrastructure.**

Government agencies need to shift to sharing hardware/network resources, similar to Web-hosting services, so that excess capacity is shared, and expertise is localized.

#### **Digital divide.**

There were five main categories that we felt would put people when dealing with the Digital Divide: income, disabilities, rural areas, and, user issues, e.g., mental health issues (not everyone is easily trainable), and age. What is the technology that needs to be "basic service" for all people? The size of the digital divide depends upon the services that we're talking about: email, Web, Word, Access, Java, etc, a person will all have different levels; what is required for what levels of participation? We are aware that some digital divide issues not necessarily difficult to attack. Solving some (possibly easy) digital divide issues may help in attacking other "social divides" that have been difficult to make progress on in the past. Attacking the digital divide in these cases is a surrogate for solving other problems.

#### **Personalization.**

In order to address this issue we need more emphasis on mass customization of data, so that the "right" information is delivered to people without cookies (privacy issues). One approach would be to deliver a large chunk of data to the client it may require bandwidth, do customization at the client (more private). Another approach would be encrypted queries (client reveals information in a way that the server can process without understanding), encrypted data (server returns data that a client can process without having full access to the database).

#### **Multilingual access.**

One major problem with the digital divide is due to the multilingual problems (not everyone speaks English). We need to make sure to have Enabling technology that has the capability to have a Cross-Lingual Information Retrieval (CLIR) i.e., Spanish queries return English documents; Automatic machine translation (MT), because it is impractical (too expensive, too slow) to translate every document into every language. The Goal is to have automatic translation for greater dissemination, but government requires great accuracy; what happens when people make decisions based on incorrect translations? How are the choices of no information (because a document isn't available in a target language) balanced against available, but possibly inaccurate information (due to translation errors)? We feel that research on government-specific issues in MT and CLIR is needed. User interface issues (how are limits of translation accuracy conveyed, how is the possibility of multiple, somewhat different translations conveyed)?

#### **Information distillation.**

In today's day and age it is now so easy to get more information than a person can (or wants) to read. Many people give up before finding what they need and people need software that will help them analyze or distill out the information that they need. We need enabling technology: information filtering at the client level; basic research is fairly far along, but little research so far on deploying in personal systems: question answering software at the server; this is an emerging research area, capable of extracting simple, factbased answers from retrieved text documents; more research is needed to handle the kinds of questions people would ask of government portals. clustering of search results (to show what's out there) and iterative retrieval (to drill down along a particular dimension). Drill-down interfaces: How do we make them easy for people to use and understand?

#### **Automatic analysis of public input.**

How does an agency handle a million user responses to a request for public feedback? The answer: text categorization to sort people into bins (by viewpoints or other characteristics); effect analysis to identify level of emotion; and, personalization of (email) responses from government officials

### Policy consequences of increased public participation.

This area seems as if it's under studied. There should be more rapid public feedback that may imply less time for deliberation. There should also be more opportunity for demagogues. As well as authentication: Is a person providing feedback from your city, state, country? How does authentication balance against anonymity (although historically the government ignores anonymous feedback)