

Workshop on Mechanisms for Improved Food Safety Data Access and Sharing

Washington, DC | March 2, 2007

Lessons Learned and Key Ideas

DISCUSSION DRAFT: March 7, 2007

NOTE: This list of ideas was drawn from discussions at the March 2 Food Safety Information Infrastructure (FSII) workshop, and is intended for solely for discussion purposes. It does not express relative priority or level of support among participants. It is not intended as a consensus document.

SESSION 1: DATABASES AND ELECTRONIC NETWORKS: INFORMATION SYSTEMS FOR IMPROVED FOOD SAFETY

PRESENTATION:

Ellen Olson, FDA/ORA -- *eLEXNET (Electronic Laboratory Exchange Network): Collaboration to Ensure Food Safety*

Notes from presentation and follow up discussion:

- Standardization and simplification of reporting is critical if there are multiple data partners
- Systems may serve multiple purposes in the FSII. For example, eLEXNET has 3 roles:
 - o Repository for sharing of food testing data between labs, including information on methods
 - o As communication platform for FERN
 - o To communicate summary information to DHS
- Role of FERN in eLEXNET is interesting example of system adaptation: when FERN was created in 2003, it gave eLEXNET increased exposure in laboratories, but functionality of eLEXNET and needs of FERN are not identical. FERN is both a benefit to eLEXNET and a challenge:
 - o FERN brings a rapidly growing user base, but needs a flexible data entry system that eLEXNET doesn't have, so this may require rebuilding of eLEXNET
 - o It's a challenge (technical and resources) to address FERN needs and also fulfill original primary purpose of the system
- Needs of a system are likely to change over time, but systems are often hard-coded and difficult and expensive to alter or adapt
 - o A common theme is that many systems are custom built at the "mercy of the contractor" and are very difficult and costly to change, and they often require continual maintenance that is also a tremendous resource issue
- There are data ownership issues with aggregated data sets: for example, in eLEXNET, you cannot publish on the data without permission from each data owner, or lab. Also, the system does not include isolates: one must contact the lab, though eLEXNET does provide contact information
- Data in eLEXNET may include information that precludes allowing broader access:
 - o Data may include proprietary information about firms or plants
 - o Data may be sensitive: FERN data is not open to non-FERN participants
 - o Data may be related to an ongoing investigation, and so linkages between isolates and a particular firm should not be used indiscriminately.

- Data in eLEXNET may be of interest to other government agencies such as CDC: it may be worthwhile to develop some data-request protocols, especially for summary data
- Analysis of aggregated data is limited due to differences in labs, which may use different methods, and purposes of original collection – need to be careful not to misinterpret information

PRESENTATION:

Tracy Ayers, CDC -- CDC's Foodborne Outbreak Reporting System (eFORS)

Notes from presentation and follow up discussion:

- Reporting of foodborne outbreaks has changed quite a bit over time, including multiple versions of eFORS and upcoming move to ORS (which includes non-foodborne outbreaks)
 - o Newer electronic systems may result in rapid increase in amount of data collected, but quality of reported data may be lower as there is more and more information for those overseeing the system to keep an eye on and assess for quality
- For reporting-based surveillance systems, standardized forms are critical, but even so, there may be differences between states or localities in how they use the forms or which fields are complete
 - o Attribution from outbreak data is difficult because foods are not standardized, so there is a lot of variation in what is in the system (e.g. “hamburger” may mean sandwich or ground beef)
 - o Due to differences across localities and with terminology, CDC cannot make raw data available; the “line listing” approach ensures consistency in data – data quality issues persist, so epidemiologists and statisticians assess quality and completeness of data
- System design is very time consuming – for the new ORS (outbreak reporting system), it took a full year just to get data needs from different offices and programs within CDC
- Changing the system is very difficult, especially due to security issues
- Changing the system may impact the ability and ways in which data may be analyzed
- System design may impede analysis; for example, the epidemiologist in charge of outbreak data does not have root access to data; only the IT people see what’s on the server
- Standardization and automation may improve data access:
 - o Standardized analysis might shorten the time and effort needed to analyze data and answer questions about outbreaks
 - o Automation may improve timeliness and allow customized reporting that satisfies specific needs - if done via website, it may free up a lot of staff time, but need to make sure users know what they are getting.
- Primary obstacle seen with improving outbreak data is resources, resources, resources
- Laws such as data quality act and paper reduction act make it difficult to improve systems because these require a very long and intensive approval process; as a result, reporting forms may not change for a decade despite ever-changing needs in reality
- For systems that collect information of interest to many different stakeholders in the public health community, it would be useful to have broader stakeholder inclusion in system design
 - o But federal agency is very limited in what kinds of information it can collect.
 - o For example, FSIS stopped collecting how much of each kind of product is made because they couldn’t show the data being used, but now when they want to go and use the data, it isn’t there. Couldn’t look down the road far enough.
- More rapid reporting by CDC might encourage greater participation by states not currently participating at a high level – some states may not feel they get much out of the system so they do not have an incentive to participate, but better and more timely reporting may provide this incentive

- By reporting more information (such as contributing factors) even if it is missing for many outbreaks, it may encourage improved reporting of these data
- Due to delays in “close out” period, summary data from one year may not be available until the middle of the following year.

PRESENTATION:

Vijay Juneja, USDA/ARS – *Predictive Microbiology Information Portal access to ComBase (Combined Database for Food Microbiology) and PMP (Pathogen Modeling Program)*

Notes from presentation and follow up discussion:

- Focus groups of potential users may be useful for identifying issues and problems at an early stage; PMIP just held a workshop of 30 targeted people to give feedback on the site, and they will have to go back and change some things based on feedback
- Systems designed for non-technical users face particular barriers – PMIP is designed for small and very small food processors, many of whom are not regular internet users and do not have education beyond high school; at first use, it was difficult for them to figure out how to use the website
- When sharing information with non-technical audience, some translation and simplification of information is key; for example, regulatory information may be presented in 4-5 pages, but small processors want it in 4-5 lines.
- Prioritization of information is also important, but difficult.
- Information systems may be useful for presenting models and data that save users the trouble of having to search or review relevant published literature or create duplicative models based on same data; ComBase is example of repository of published models; PMP allows users to enter relevant numbers into easy-to-use models and makes sure numbers entered are in the approved range
 - o Repositories may provide access to information only summarized in publications
- Although ARS makes PMP available, USDA is not responsible for predictions, and firms can only use these predictions as guidelines
- PMP and ComBase are intended for specific audiences, but are open to everyone
- Systems sometimes exist in stovepipes and result in low awareness of data and models; many have not heard of PMP or ComBase, despite that it has been being used by regulators for decades
- Users of systems may not be anticipated; models in PMP may be used by economists attempting to figure out costs and benefits of a regulation, but this is a far cry from a small processor attempting to validate a food safety management system
- In system with wide and potentially uninformed users, need to be careful how data are interpreted; first page of PMP is a disclaimer, and ARS recommends users to use most conservative assumptions
- Version control may be a concern; for systems that are run on users own machines (that is they are downloaded and run, it is difficult to ensure users are using up to date data
- There are quality control issues with “open” systems that allow for anyone to upload; ComBase has developed some software controls to catch outliers and to screen data, and it may take months between when model is uploaded and it is made available to users
- Expensive systems may be worth the investment – it is estimated that over 15+ years, the models in PMP may have cost between \$2 and \$25 million, but the cost savings to the economy have been incredible, and this investment really spawned the field of microbial risk assessment.
- Users can’t just run with the data, they must use their own ability to judge quality – the role of the system is to give access to data, not to validate each data point; systems like GeneBank and ComBase make data and models available, but there may be outliers or errors in there

SESSION 2: FOOD SAFETY WEBSITES AND LISTSERVS: USING THE INTERNET TO SHARE AND DISSEMINATE INFORMATION

PRESENTATION:

Steven Gendel, FDA/CFSAN -- *Foodrisk.org: The Online Resource for Food Safety Risk Analysis (formerly the JIFSAN Food Safety Risk Analysis Clearinghouse)*

Notes from presentation and follow up discussion:

- Standardized vocabulary is important for proper use of data; FoodRisk.org includes a thesaurus to define terms used on the website
- Website may serve to build a community by serving as a centralized information location
- The type of data a system includes reflects domain of study; Risk analysis is by nature a “meta-domain” – it sits over a large number of disciplines; therefore, FoodRisk.org captures a wide variety of information – risk assessments, models, data, tutorials, educational material, links, powerpoint presentations from food risk analysis events, etc
- Data repositories may be valuable for otherwise unpublished data that may nonetheless be incredibly valuable to regulators or risk assessors. Negative data or simple baseline collection may be difficult to publish in a journal, but a data repository such as FoodRisk.org may make that data available
- Tracking is difficult, especially if the website doesn’t “own” the data; the website may request users to cite in publications, but even this is difficult to track
- Citations to websites and repositories may serve to measure utility of site, as well as to promote site to others; for data “exclusive” to FoodRisk.org, the site requests users to cite the site in their references
- A website is implicitly international; although we create something for Americans, it may prove very useful to others; FoodRisk.org is used by many in Japan, Korea, etc.
- There are at least two benefits of associating a data system with a university instead of making it solely a government run initiative: (1) The system may avoid burdensome limitations of government data regulation; FoodRisk.org is a project of JIFSAN, and website is based at UM, so many government rules about data do not apply; (2) Students may be able to work on the system for free or very low fees; a lot of the work that has gone into FoodRisk.org has been done by computer science students who wanted to try new things
- Flexibility of the conceptual idea is important; how can you build in the capacity to grow in the future without it being painful and expensive? On the other side, flexibility is something that may be impossible if every iteration goes through a lengthy approval process.
- Learn to embrace chaos; There is just so much information out there that is relevant, if you try to standardize too much, you will lose some things in favor of simplification
 - o Systems evolve very rapidly; the rate of growth in information systems and computer capacity is breakneck, and we need to make our websites and information systems be able to handle these changes
- Third-party data blinding is a concern; although a site may be open to industry putting up their data, it is impossible to do so without a lot of effort to blind the data
- Copyright concerns; who owns what is up on a website or how the information is presented?

PRESENTATION:

Yvette Alonso, USDA/ARS/NAL – *The Food Safety Information Center: “Providing food safety information to educators, industry, researchers and the general public”*

Notes from presentation and follow up discussion:

- Incentives are critical for data repositories, otherwise parties may not use the system
 - o NAL is mandated to collect information on research across the federal government and within industry, but other agencies are not mandated to provide this information, and private industry has no interest. Agencies such as NIH don't see any benefit of providing information to FSIC – “why should we work with you?” – especially if the other end may need to devote resources and training to do so.
 - o Funding for FSIC comes through ARS, so collection of information about ARS research is better than non-ARS research
- Standardization issues arise when collecting information from multiple agencies, which may use different structures of information or might use different terminology
- Resources are a huge problem: FSIC's budget has remained fixed for fifteen years, but our responsibilities have increased; salaries and I.T. costs rise – what gets pinched is most often the technology
- Websites may be affected by decisions from above that affect utility of the website: In 2006, USDA mandated all websites had to have a standardized look, but provided no funds for this conversion, and FSIC had to abandon certain initiatives
- Due to regulations, foodsafety.gov cannot provide links outside the government, whereas the NAL is a library, and can link to third party websites
- It is almost impossible for the government to stay up to speed with changing technologies; systems get locked in and it takes so much effort to change them
- Providing access to information may mean reformatting it for search engines; NAL is putting major catalogues into XML so Google can find it in its searches.
- Government agencies should keep an eye on new technology; RSS feeds may be a simple and useful technology that allows regular users to keep track of site updates, such as when data is made available, or if there is a fast-moving story such as an outbreak
- Given complexities and rapid changes in system, need to help people search
 - o List of “best” links according to expert judgment – going through Google results for a search to pick out (via a set process) which are of the highest quality
 - o Federated searching – searching a group of websites or a sub-set of all available pages on the internet
 - o How can we use Google to improve food safety data searching?
- Focus groups are useful to really validate a system; FSIC is going to perform focus groups to see which of its efforts are most useful, and which need the most work
- Despite automation, performance of systems is often based on ability of people talking to one another and establishing relationships; the human component of these systems is affected greatly by job turnover or constant reorganization

PRESENTATION:

Doug Powell, Kansas State University -- “I Do News” - The Food Safety Network

Notes from presentation and follow up discussion:

- Copyright can be a huge issue; who owns the information or the way it is presented? FSN moved from U Guelph to KSU, but everything is copyrighted to Doug Powell not the University
- News stories can provide a lot of historical information: example, when peanut-Salmonella outbreak hit, FSN immediately found past stories related to peanut outbreaks that provided background information in minutes
- Need to embrace more “push” technologies, and less “pull” technologies; FSN is successful because it puts the news in people’s inboxes every day
- Simply putting information out there isn’t enough, we have to put a lot of effort to get into people’s hands, and from there to get it into their heads
- Know the audience of information and cater to it; most of the people working in kitchens are teenagers and many are non-English speakers
- To reach people, you need to make information “compelling:” we need to embrace pop culture and humor; we can still make information science-based, but we need people to read it, so we need to make it interesting and we need to speak their languages
 - o How do you keep people interested when there isn’t an outbreak? Use entertainment, but need to be rapid and reliable and scientifically sound.
 - o We should watch what the activists do – their information may be lousy, but they know how to capture attention.
- We don’t need to remake the wheel, we don’t need every local health unit to recreate hand washing signs; we need more people assessing whether the signs are helping.
- Need to answer: How can you assess true impact of communication?
- We need to take advantage of observational and behavioral research to really understand what people do with information
- In a fast news culture, turnaround is key; waiting for annual reports is not going to work
- Need to stay on top of new technologies and opportunities, and simply see what sticks

PRESENTATION:

William Krueger, Minnesota Dept of Agriculture -- FoodSHIELD: Meeting the Challenge

Notes from presentation and follow up discussion:

- A single website may serve as a portal to a great number of data sources
- Standardizing the look and tools of a website means that people can become used to a single system
- Integration vs. linking – hard to sort through links, & integrating data in one place makes it easier to find (“FoodSHIELD is the Christmas tree, and each dataset is a bulb you can hang from it.”)
- Standardization of job roles and responsibilities is key because of differences between states and communities; by standardizing job responsibilities, you can find who you are looking for by the information you are looking for, not by going through organizational structures to try to find the right person
- An integrated web site can help overcome FOIA and turf issues by achieving collective work towards the same solution
- How can we leverage technology to ensure accurate contact info? Detailed standardized reporting and active attempts by website staff to ensure that each laboratory is providing the most up to date information to FoodSHIELD on a regular basis

- The network of human contacts is the most fundamental network for information sharing among food safety personnel and is the intuitive starting place for finding food safety data;
 - o The core of FoodSHIELD is the contact management system, but the system can carry any sort of information, whether data itself, training software, videos, etc, that radiate outward from the directories
 - o Related to this, FoodSHIELD also captures equipment by model and manufacturer to create a national inventory; can use this information to find experts on certain techniques
- Incentives for participation: grassroots efforts to keep everyone involved in design, plus peer pressure, plus using associations to make states participate
- Access rights are difficult in area of food defense as the amount of sensitive information and clearance requirements increase

SESSION 3: DISCUSSION

The Single Web Portal: Putting aside who would do it, should we aspire to a food safety website that can serve as the starting point, a place where everyone would start and then fan out from there?

- Is there another example of this out there? Wikipedia?
- If you start with a top-down view, it will never work; need to stick to a grassroots approach – we all speak to different people and have different niches
- There are institutional concerns when one party’s site becomes embedded in another’s site
- There are copyright issues
- There are large perils of such centralization – it may be more vulnerable to electronic attacks
- Centralization takes away from variety and diversity of information and viewpoints
- A portal may serve to increase connectivity - perhaps rather than embedding information or making things available through “sub-sites”, linking is better because it maintains original authorship, but increases visibility
- Whether it is a single site or not, we need to overcome the silo effect
- Upside is to perhaps find new audience
- Would hate to see hits go down because of it; it’s how we’re measured, for better or worse.
- Two counter examples
 - o GeneBank was created as the master list of DNA/protein sequences, but there are hundreds of other systems devoted to this, even if GeneBank is the main one – you can’t mandate something like a single point
 - o Yahoo: Created as a directory, but as the web grows, this becomes increasingly difficult; this kind of organization has been replaced by Google searching approach
- JIFSR was going to be such a system, where people could support each other, interact, talk, share documents, but it never came to be; who knows whether it would have been useful
- It’s analogous to single food agency; we may say let’s all do it this way, but we may be quite incompatible; the benefits of that kind of organization may be overstated
- The real goal should be to keep up with our clientele; who are they, what do they need, what new technologies can help; as we all have different audiences, it’s likely that each system moves in a different direction over time
- Diversity is critical, and just as for a strong environment we want rich biodiversity, we need rich “infodiversity”; what we do need to do is promote the flow of information through the system
- Silos are good because they do promote a wide variety of views, but the key is knowing what silos are out there other than your own

- The biggest problem is the lack of a common terminology; we can't standardize or normalize, but it is a huge process requiring a lot of human energy and input to build up this common language
- So the big-picture aspirational vision may not be a single portal but niched portals that are linked; the central driver is that there is additional value to the work that everyone does, or the data or information collected, that may be valuable to the system as a whole, or to others in the system, so the vision is something that helps facilitate that flow of information

Institutionally, where does the investment come from?

- At the end of the day, you have limited resources to reach out to other groups, we have gotten so thin, we can hardly get done the bare minimum
- Any number of agencies and programs, if the group of people who use the tools can get together and discuss how to improve things, reduce redundancy, etc. Even getting together today is a start.
- We need a consortia focused on database systems to have these discussions more regularly; getting people together to talk about database sharing and maintenance wouldn't cost much, and it may be easier to obtain funding if we work as a team
- There is the idea out there of build-once, reuse-several-times; for example, we built the library of methods, which then we make available to others. Doing this can help reduce redundancy
- The real trouble is that the benefits accrue to the system as a whole, but all the resources and effort have to come from individual institutions that may not get enough benefits to warrant their costs.
- Many other disciplines or groups have professional associations that wrestle with these issues; if the topic were more specialized (e.g. microbial methods development) the answers may come easier, but given the breadth of disciplines and number of institutions, it's unclear who takes the lead.

Given the diversity in the system, nobody can "run the FSII" but what can we do to make the system easier to get your arms around? How can we overcome the lack of a single responsible institution?

- We need to bring in experts from informatics.
- There is a revolution going on with information management: web 2.0 and so on. How can this technology be used to help us? Technologies that embrace the dynamic changes, and create dynamic websites or feeds or email alerts are something we should look to tap into. We may be able to use these automated technologies to reduce the silo effect.
- The government is going to lag behind, but we need to find a way to drag the government into the 21st century by using as many different vehicles as possible
- We need to move towards open source approaches; there is a lot of technology that is free and increasingly robust that can be customized

What are your last words on the subject? Let's go around the room for final suggestions and ideas.

(Responses have been roughly organized into some composite themes, in no particular order):

- Diversity is important:
 - o Embrace diversity (repeated by multiple people)
 - o Diversity is key, as it promotes innovation
 - o Important to keep diversity of views and diversity of entry points in system
 - o Don't centralize the system – diversity is healthy
- Improve standardization:
 - o Work on standards (repeated by multiple people)
 - o Think about standards; need to be cognizant of standardization
 - o Standardization of technology so we can look at data across data sets

- Standardization is daunting but we need to move forward piece by piece
- Food safety community should join the consortium for establishing public health standards
- Consider users and potential users of data and systems:
 - Keep in mind that there are a variety of potential users of information
 - Remember your audience, especially at the local level
 - Important to know who your user is, but look beyond your typical user; a lot of info today would be useful to my members – most data sharing is between researchers, not to others in industry or private sector
 - Use focus groups - Keep open mind as to who might want to use information
 - Need to distinguish between those in the food safety community and everyone else; what works for one will not likely work for the other
 - A lot of shared information is only available to government, but it needs to be more consumer or user friendly to have a real impact; let's work ourselves out of our jobs.
 - Build flexibility into your system and between systems, especially with respect to how users interact with the system
- Evaluating impact and honing communication strategies:
 - Remember why you are doing it and find a way to evaluate it
 - The question that drives me is: how do I take this information and make it useful on the streets? Many of those that most need this information do not have internet access
 - How do you reach people? For example, in the peanut butter example, people would call up and ask questions already answered in the press release or on the website
 - Need to discuss how we measure impact – what has worked and what hasn't? What are the model practices?
 - The challenge is: how do we get a message out that they listen to and respond to?
 - Need to learn how to measure impact of communication messages.
 - Measure effectiveness of efforts – if we can show that we are saving lives, we can get money for our systems.
- Increase connectivity in the system:
 - The biggest problem is awareness: how do you make people aware of these things?
 - Having one place where I can see all the websites would save me lots of time on Google
 - Partner and “plagiarize”; single data entry point rather than having to enter the same information multiple times
 - Need to find ways to cut across stove pipes; diversity is good, but without improved connectivity between systems, it is nearly impossible to know what is out there; you have to know something exists before you can search for it
- We need to come together and address issues collectively
 - Perhaps we should have a forum every couple of years to keep ourselves informed about what others are doing
 - Partnership is key, as the community can take itself where it needs to go. “Actively managed community of practice” is a concept where a facilitator helps to lead the group towards moving towards agreement on what has to happen
 - Look into partnering with other federal agencies in data-sharing; for example, I have isolates and PFGE patterns from plants, but I'd like to know if this has been found in outbreak data, there is no common nomenclature.
 - Technology can't do everything, so we're going to have to work together to find partnerships.
 - Hope this is the first meeting like this of many.

- We all have unique talents; working groups to facilitate diversity and use our strengths
- Make sure industry is more involved moving forward
- Acknowledge and address resource needs:
 - Be patient: it is not that we are intentionally withholding our data, we are just overworked
 - Resources are slim, but if we put collective voices together we can make greater impact and show that this issue needs funding; need to find a way to keep food safety on the radar in absence of an outbreak
- Embrace and exploit technologies:
 - Be smart with data and technology: where is web 2.0 taking us? “Self-aware data” - data is wrapped in an envelope of description, the data tells the user how to use it
 - Move to open source
 - Technology that allows you to do a lot more w/ less. How can you funnel and structure to get to your users?
 - Simple things can improve the current situation: some agencies have done a great job letting people subscribe to listservs that share info, but others don’t. FDA puts out good guidance documents, but I can’t get them unless I go digging around their site looking for them.
 - Wikipedia is the future
 - Be creative: a group of students at Harvard takes Google maps and matches it to different live feeds so you can watch an outbreak on a map in real time.
 - Citizendium – an outgrowth of Wikipedia developed by Larry Sanger – an “expert guided” version intended for government and civic information, but which might be more applicable to food safety information than the free-form wikipedia
- Accessibility to existing data:
 - The only way to improve data “sharing” is if sharing is embraced by the holders of that data as a key responsibility – a lot of the time the collection of the data is the REAL end point and we don’t put much effort on the follow through.
 - If you don’t build it someone else will: For example, CSPI is making a searchable online outbreak database, and this will likely spur CDC to do the same
 - Just as we are making restaurant inspection data more available, I see other information going this same route.
- Gaps and research needs:
 - Food defense is a huge gap in the system
 - Research in food safety informatics
 - A lot of the databases aren’t that useful at the local public health level; rather, need consumer education, behavior modification in these lists of websites. Need more data on behavior and what works, more sharing of lessons pertaining to behavior and whether training and certification makes a difference
 - Need more data on educational research – who would a person listen to? Etc...

Compiled from notes by Marin Schweizer, Michael Batz, Jan Powell, and Mike Taylor