**Scientific computing workshop** position paper for ECSCW 2003

# Managing Information for Mass Fatality Identification: Gene Code Forensics and the World Trade Center Disaster

Debra Cash and Howard Cash, Gene Codes Forensics

In late September, 2001, Gene Codes Forensics, a bioinformatics company based in Ann Arbor, Michigan, took on the challenge of creating a software application that could track the remains, personal effects, kinship data and DNA of the victims of the World Trade Center attack in order to identify the victims. This has been the most ambitious forensics identification effort in history.

Gene Codes established five goals:
Identify individual remains
Reunify partial remains so that they can be returned to families
Collect and warehouse meta-data for administrative review of reference samples (antemortem victim materials, such as toothbrushes, razors etc.)
Track samples among collaborating laboratories
Create an information management system to report metrics and make problem resolution proposals to supervisors at the New York City Office of Chief Medical Examiner (OCME)

The programs available to the OCME at the time of the disaster (primarily CoDIS, the Federal Bureau of Investigation's Combined DNA Index System) had been designed around unique identifiers, such as fingerprints, and "clinical draws" in which DNA information from bone marrow, blood samples or cheek swabs was unambiguous. CoDIS had no way to group and collapse data from dozens of fragmented remains (in one unfortunate case, a victim was fragmented into more than 200 pieces) or perform "all against all" matches. Nor could it accommodate data from degraded DNA samples (as the world knows, "Ground Zero" was on fire for over three months) and incomplete genetic profiles. CoDIS was also inadequate to deal with DNA profiles generated by commingled remains, crushed together under the weight of the towers' collapse.

Gene Codes is best known for its market-leading DNA sequencing product, Sequencher. (See HYPERLINK http://www.genecodes.com www.genecodes.com for more details.) That product was designed with a profound commitment to user-centered design and an understanding the work practices of bench scientists in a number disciplines. The OCME, as well as the US Army and other government agencies, were already customers of Gene Codes on September 11, 2001, and turned to the company for support within days of the disaster.

Gene Codes Forensics embarked on a complete—and difficult—ethnographically- informed workplace mapping and inventory assessment, tracking how information flowed among bench scientists, work groups, computer systems (such as installed laboratory management systems) and participating labs, how information was tracked and transformed at each state, and how it was reconciled at the point of identification and confirmed before identifications were released to medical-legal investigators and the victims' families.

There was special concern with understanding how information gathered—and mistakes made--at one point in the process could have deleterious efforts on the ease or accuracy of identification later in the process. (As a simple example, in the first two weeks after the WTC disaster, the New York State Police or other subcontracted agencies received over 12,000 individual items, such as toothbrushes, razors and hairbrushes. This astonishing flow of materials overwhelmed the processes that had been put in place to manage it on an emergency basis. Later, an "administrative review" process—primarily a paper-based research process--would have to be put in place to ascertain that "John Doe's" toothbrush actually *belonged* to John Doe and had not been donated by a the family of another missing person.)

M-FISys (pronounced "Emphasis"), the product developed for the WTC identification effort, had to address—and resolve--a number of issues of special importance to CSCW researchers, especially those working in the biological sciences. These include:

M-FISys had to be a "tool for skilled work." The software application itself does not "make" identifications: only certified forensic scientists with a certain level of expertise may make identifications and approve the issuing of death certificates.

M-FISys had to display and integrate information coming from a number of sources: the medical examiner's office, the state police forensic lab which was handling personal effects and family references, plus a number of high-throughput commercial laboratories. On September 11, 2001 these organizations not only had incompatible *applications*, they had incompatible *networks* and in most cases, incompatible and often confounding nomenclature schemes that only became more baroque over time.

M-FISys had to allow the forensic scientists to interrogate the raw data *behind* genetic profiles generated by the participating labs and reported in the system database

MFISys had to build in algorithms to encode legally-designated kinship likelihood ratios, an exemplar of instantiating or "pointing to" extrinsically defined scientific standards in any scientific practice

M-FISys was developed *iteratively* in tandem with rapidly changing work conditions and scientific practice. Although the particular group of end-users (staff criminalists) in the Medical Examiner's office remained relatively stable, their activities changed, in some ways dramatically, over time, in some cases *because* the M-FISys software offered them new capabilities. Using Extreme Programming (XP) methodologies, where programmers work in pairs and unit tests and acceptance tests are written before new functionality can be added, the Gene Codes Forensics engineers deployed new releases of M-FISys on a *weekly* basis, starting in December 2001. (There have been more than 70 releases of the product as of June 2003.) New functionality was added after observation and direct negotiation with users about

their most pressing priorities for any given period of time. In addition, the use of Single Nucleotide Polymorphism (SNP) from nuclear DNA, while accepted in genetic research, had never been used in forensic identification efforts. It is being added to the arsenal of identification modalities in the hope that this will help identify persons whose remains are otherwise unidentifiable. (As I write this, I believe the use of SNPs has not yet been certified for WTC identification by the national agencies involved; this information, however, has already been referred to in media reports).

The development of M-FISys is a compelling case study of a number of key CSCW concerns:

How do we design and build computer systems under conditions of incomplete information and changing workplace conditions?

How can systems be developed quickly in settings where *there cannot be a single mistake?*

What are the key differences between building a system for an identified group of beta users or "early adopters" (as are most of the cases of systems built under academic research efforts) and commercial development (such as Gene Codes' Sequencher product)?

What, if anything, are the salient differences between designing for users who will employ a CSCW system under day to day conditions (here, making identifications) and those who will be rolling up cumulative or overview data in report format (such as the lab managers reporting to the mayor and other officials)?

What, if anything, are the design and technical concerns associated with systems whose *deliverables* (in this case, confirmed identifications of victims) will ultimately be delivered to people other than the end-users (in this case, the families of victims, rather than the forensic scientists themselves)? Does this have any implications for understanding large-scale systems such as applications that will have broad public policy implications in the areas of environmental sciences, epidemiology, or bioengineering?

The proposed participants for this ECSCW workshop are

**Howard Cash,** Founder and President of Gene Codes and its wholly owned subsidiary, Gene Codes Forensics. Howard founded Gene Codes in 1988. Within three years, the company had pioneered "Sequencher" which is now the dominant desktop sequencing software product and is sold in 45 countries, commanding a 70% market share in DNA sequencing by research labs, drug companies, and the like. In September 2001, the company agreed to develop the product which became M-FISys, the application described above. Howard is the recipient of a number of important awards, including MIT New Enterprise Forum Entrepreneur of the Year, Ernst and Young's "Entrepreneur Award" for Eastern Michigan, and "Person of the Year" voted on by the readers of Genome Technology Magazine in 2002. He is serving a three-year term on the HUGO Council, the managing body of the international Human Genome Organization. In 2003, Gene Codes was nominated by the Governor of Michigan for the National Medal of Technology, the U.S.'s highest honor for technological innovation.

**Debra Cash** has worked as a workplace ethnographer and user-centered design researcher since the mid-1980s, and has been principal of New Century Enterprises, an international consultancy based in Boston, since 1991. Her clients have included Lotus/IBM, Motorola, Xerox, Zurich Financial Services Group (subcontracting to the Doblin Group in Chicago) and the Netherlands Design Institute. A number of her co-authored papers have become standard CSCW and participatory design references. In 2001, she agreed to join Gene Codes and her brother Howard to bring her analytical skills to the WTC identification effort. She continues to serve as a peer reviewer for *Communications of the ACM* as well as other journals, edited volumes and professional conferences including CPSR's Participatory Design Conference. Debra received her Master in Design Studies (MDesS) degree from the Harvard Graduate School of Design's Advanced Independent Study Program in June 1995, where she held the IDRC Graduate Research Fellowship.

**<u>Papers written or submitted on Gene Codes Forensics and the WTC effort</u>**
[Please note that papers on this topic were embargoed until late in 2002 and others will be forthcoming.]

Brenner, C.H. and Weir, B.S., "Issues and Strategies in the DNA Identification of World Trade Center Victims" in Theoretical Population Biology, 63 (2003) 173-178

Cash, D., et al "Homicide x2,793: Transforming Forensics Practice into Mass Fatality Identification", under review for 2003 conference in the U.S.

Cash, H.D., Hoyle, J.W. and Sutton, A.J., "Development under Extreme Conditions: Forensic Bioinformatics in the Wake of the World Trade Center Disaster," Pacific Symposium on BioComputing, January, 2003

Hennessey, Mike "World Trade Center DNA Identifications: The Administrative Review Process," Promega, 2002