

## Information Technology

*Information technology (IT) can be one of the most overlooked aspects of a mass fatality incident response. This is understandable—after all, most senior laboratory managers are forensic scientists, not computer scientists. However, advance planning on integrating IT throughout a mass fatality DNA identification effort saves time, speeds identification, and improves the reliability of the testing. It is crucial that the project manager include IT personnel in decisions regarding sample tracking and other business processes.*

**T**he archiving and management of the vast amount of data involved in a DNA-based identification of mass disaster victims is an enormous challenge. Because data must be retrieved, compared, and integrated reliably and efficiently, it is crucial to have sophisticated software.

In June 2005, the National Institute of Justice published *Mass Fatality Incidents: A Guide for Human Forensic Identification* (NCJ 199758; [www.ojp.usdoj.gov/nij/pubs-sum/199758.htm](http://www.ojp.usdoj.gov/nij/pubs-sum/199758.htm)). The study that resulted in the guide found that:

The process of accumulating, reviewing, and interpreting DNA data is the most challenging step when employing DNA technology to identify mass fatality victims.

Ideally, an experienced IT laboratory staff member should be involved in the management of a mass fatality DNA identification effort. IT should be the cornerstone of quality control, and the IT department should continually be searching for ways to improve work processes and turnaround time.

For example, one way to increase productivity in a mass fatality identification effort is to have the Laboratory Information Management System (LIMS) produce a daily progress report for the media and elected officials. It is preferable to develop this capability before an incident occurs, because it is extremely difficult to achieve this level of IT sophistication in the midst of a mass fatality response.

Exhibit 14 shows different functions that an IT system can support in a DNA laboratory. The arrows follow the basic flow of samples and data. In highly automated laboratories, these procedures will be monitored or controlled through the LIMS.

Regardless of their level of automation, all laboratories employ these systems, in some form or another, during routine casework and offender processing. To effectively support a mass fatality incident response, however, some of these systems require special features.

Prior to 9/11, mass disasters in the United States were relatively small in scale, allowing simple spreadsheet approaches to be sufficient for data management. However, even small-scale disasters require scrupulous data management. Although software programs exist for data management, access, and statistical analyses, the magnitude of the World Trade Center (WTC) disaster demanded enhanced capabilities.

**Our chief lesson learned in this arena is that, without validated, well-documented software programs to associate profiles from tens of thousands of remains with scores of direct and indirect reference samples, the matching process is untenable. The midst of a victim identification project is a difficult time to be beta-testing new versions of software.**

*Stephen Sherry*

**In the early stages of the World Trade Center identification effort—when the results of analyses were just beginning to arrive—we had to get information about the origins of any sample by querying several different computer systems; we often had to review paper records and ask for help from the New York State Police. The problem could only get worse, and we knew there had to be a better way to include functions in the LIMS beyond just tracking laboratory reagents and samples.**

*Elizabeth Pugh*

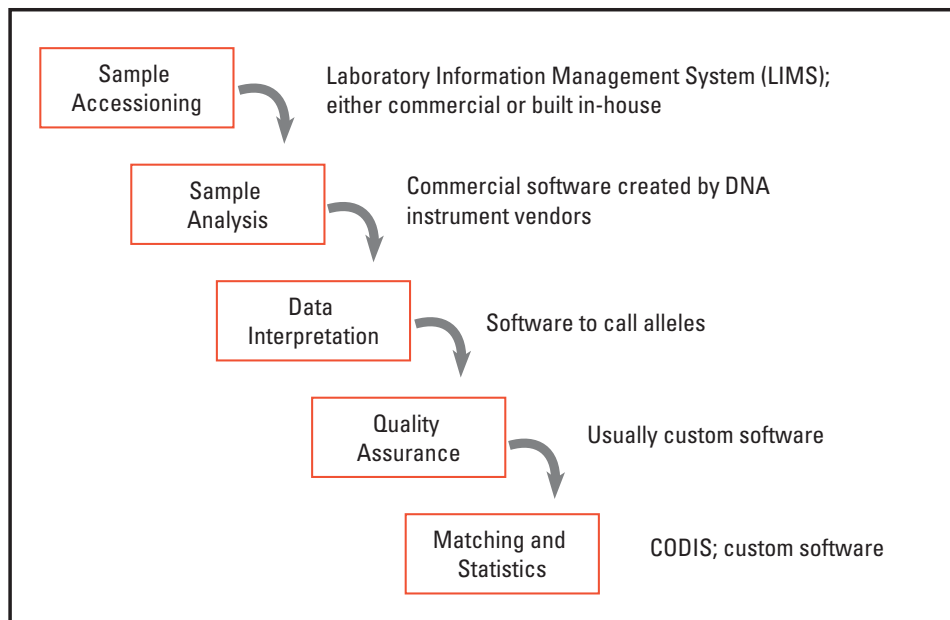
Data-handling systems are needed to integrate any customized software as well as to provide a middleware system for connection and integration of different software components. Computer software must be able to assist with many functions. It must:

- Organize, store, and retrieve diverse data.
- Integrate different software systems.
- Allow technical and administrative review of data.
- Allow for annotation and recording of problems and resolutions.
- Report metrics.
- Track samples among partner laboratories.

- Prioritize sample selection and review.
- Generate family pedigrees and calculate likelihood ratios for hypothesized kinships.
- Combine remains with the same profile to facilitate searching.
- Enable profile comparisons and statistical calculations.
- Allow for users to interact with the interpretation and evaluation of ambiguities.
- Be reasonably user-friendly.

In the WTC identification effort, the Office of the Chief Medical Examiner (OCME) contracted with a private vendor that developed software with the above listed requirements in mind. However, because the software was developed in the midst of the identification effort, and was not previously documented and validated “shelfware,” the majority of the kinship analyses and (initially) the remains matching had to be conducted using several commercially available programs. These programs were supplemented with customized patches developed by Kinship and Data Analysis

### Exhibit 14: Information Technology in a DNA Laboratory



Panel (KADAP) members who were deployed to the OCME. Without such software, the success of the WTC identification effort—nearly 1,600 identifications made and nearly 20,000 remains profiled—would not have been possible.

Another software tool used in the WTC identification project was from the FBI's Combined DNA Index System (CODIS). Two of CODIS' four files—the Missing Persons and Unidentified Human Remains Index (CODISmp) and the Reference Samples from Personal Items and Family Index—allow the search of DNA profiles. The use of mtDNA profiles as a screening system is facilitated by the introduction of the CODISmp system. Although designed for missing persons, the system may be used to search for DNA profiles of mass disaster victims.

## Sample Accessioning/LIMS Requirements

All LIMS products have a sample accessioning capability, usually centered on a case number. Each case number has multiple items, or submissions, associated with it. Some LIMS may print a barcode that assists in chain-of-custody documentation and the creation of management or status reports.

In a mass fatality incident response, human remains, personal items, and kinship samples must be accessioned. Laboratories usually retrofit these special requirements into their existing LIMS. There are a number of benefits to this approach, including that:

- Laboratory personnel are familiar with the sample accessioning process and can avoid the learning curve associated with new software.
- Chain of custody is documented and controlled using tried-and-true processes already in place.
- There is no need to purchase additional software.

The typical strategy for accessioning human remains is to assign all fragment submissions to a single case number. This is relatively straightforward if the LIMS allows a single case to have thousands of submissions (one for each remains sample). If the laboratory's LIMS does not allow

for a large number of samples to be associated with a case, the laboratory will want to consider developing a system to link the cases so that all samples can be associated to each other and to the identification effort.

The commingling of remains presents another problem. For example, after a remains sample is accessioned and analyzed, the laboratory may discover that it belongs to two or more individuals. The DNA may show that the bone and tissue come from different donors, as happened in the WTC attacks, where remains were severely compacted.

From the moment commingling is discovered, the laboratory will have to assign a new submission number to one of the items, then track both items separately. This principle would apply even if there are more than two profiles from a single sample—each profile would require a new submission number. Some LIMS systems may allow a sample to have multiple DNA profiles; regardless, both samples will have to share the initial chain-of-custody and accessioning information.

Several other scenarios may further confound sample tracking. For example, there may be multiple victims from the same family, in which case the situation is complicated by partial profiles with overlapping genotypes or by full or nearly full DNA profiles from remains that are needed as reference samples for a related victim. Such difficult situations can occur and must be accommodated.

The typical strategy for accessioning reference samples is to assign each victim a case number and add reference samples as submissions under the case. The case number is important because it represents the victim's family and is used to group personal items and kinship samples for kinship matching.

Assigning case numbers is not a complicated issue in a "closed" incident—for example, when a flight manifest contains names and addresses that can be tracked. During reference sample collection in a closed event, the family assistance center can review the list of victims and assign reference samples to the correct case number. In instances where victims have similar names, the family assistance center can ask family members for clarification during the collection process.

However, assigning case numbers in an “open” incident is much more complicated, and may tax the capabilities of the laboratory’s LIMS. Because there is no definitive list of victims in an open incident situation, the family assistance center—not knowing, for example, if there is more than one victim named John Smith—cannot simply assign case numbers to victims. This problem is exacerbated when reference samples are collected in an open-house forum, where members of the same family visit the collection center at different times. This also can lead to errors in the collection process, including variations of a victim’s name and perhaps even date of birth.

Therefore, during accessioning, staff entering data should avoid the temptation to reconcile name variations. Rather, data should be entered exactly as specified on the collection form. Any necessary case number or victim reconciliation should occur after the final list of victims is established. This

approach to accessioning will generate more case numbers than victims, but it will preserve all the information provided by the donors.

Unless care is taken when identifying and assigning case numbers to the potential victims, the laboratory will be forced to reconcile originally assigned case numbers with a later, more refined list of case numbers. Some of the originally assigned case numbers might have to be divided, and others might have to

be consolidated. The most important part of a process that requires a regrouping of reference samples is preserving the original case number so that:

- Samples do not have to be barcoded again.
- New case numbers do not have to be issued to families.
- The chain of custody is maintained.

Exhibit 15 presents different scenarios of reconciling case numbers with victims. It is important to keep in mind that some LIMS products may not allow reconciliation of case numbers with victims.

Exhibit 16 presents some additional capabilities that require LIMS support.

If possible, the software used by the family assistance center to collect reference samples should interface with the laboratory’s LIMS. This avoids duplicate data entry and eliminates the potential for data-entry errors. At a minimum, the two systems should have compatible barcodes so that the samples do not have to be barcoded again during accessioning.

Ideally, the laboratory’s LIMS will be able to:

- Store the data included in the reference sample collection forms (see appendixes B, C and D to this report).
- Capture photographs of remains samples and personal items and digital images of handwritten collection forms.
- Store family pedigrees and allow a victim sample to be used as a reference sample for another victim, if necessary.
- Allow cases to be divided and combined.
- Track samples to and from multiple laboratories.
- Track multiple testing of the same sample.
- Alert the end user to discrepancies in data.
- Prioritize sample testing and data analysis.

## Quality Control Software

Software is not only a case-tracking tool. It is also a critical component of a DNA laboratory’s quality assurance and quality control programs.

Quality metrics collected and tracked through software are used to refine and improve the laboratory’s quality assurance plan, and software tools often are employed as quality control mechanisms. Mass fatality incident responses have several, specific quality control needs:

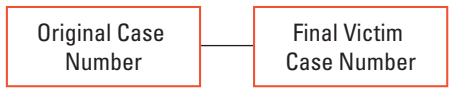
- **Identify conflicting STR results.** Remains samples and personal items may not yield usable DNA profiles on the first analysis attempt. The laboratory may choose to reanalyze these samples under altered conditions in the hope of producing a complete—or a more complete—profile. The laboratory will need to compare the results from each analysis to identify and resolve conflicts.

**We found that instituting quality checks throughout the identification process ultimately saved time and effort. By continually validating the accuracy of the data and results at each step in the analysis, we could identify potential issues before they became impediments to an identification.**

*Amanda Sozer*

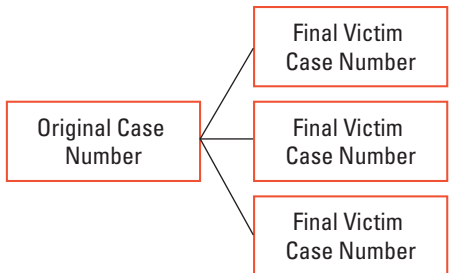
**Exhibit 15: Reconciling Case Numbers With Victims**

Case 1: An originally assigned case number maps directly to a victim.



No additional work required.

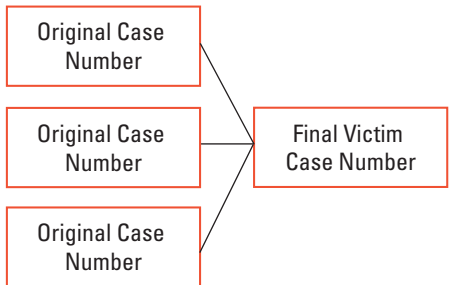
Case 2: An originally assigned case number contains reference samples for multiple victims.



This can occur when multiple victims have the same name or reconciliation is attempted during accessioning.

More difficult to resolve from an IT perspective than Case 3.

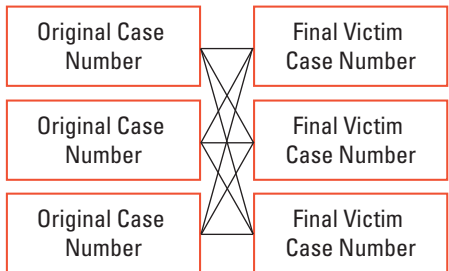
Case 3: References for several originally assigned cases actually belong to one victim.



This can occur when family members report multiple variations of the victim's name.

Relatively simple to resolve from an IT perspective.

Case 4: Combination of Cases 2 and 3.



This can occur when victims have the same name and family members report variations of the name (e.g., mixing up Jr.'s and Sr.'s).

Most difficult to resolve from an IT perspective.

Case 5: Orphaned original case number.



Reference samples were collected for a nonexistent victim. This can happen in "open" incidents.

Case 6: Orphaned victim.



The victim has no reference samples. Identification can be made only in a "closed" incident.

**Exhibit 16: Additional Capabilities Requiring Laboratory Information Management System Support**

LIMS Feature	Description
Support multiple DNA technologies.	Some LIMS products actually store the DNA profile associated with the sample. If the response employs several DNA technologies, the LIMS must support the various profile types (e.g., STRs, mtDNA, single nucleotide polymorphisms (SNPs)).
Allow samples to be tracked on multiple microtiter plates.	One sample may have numerous extracts. Depending on the types of DNA technology conducted, a sample may appear on multiple microtiter plates (e.g., STR plate, mtDNA plate, SNP plate, various reextraction plates). This feature is particularly important if extracts are shipped to partner laboratories.
Allow one sample to have multiple sample numbers.	<p>Each sample in a mass fatality response will have several different sample numbers, each assigned during a particular business process. These sample numbers are actually references into other databases.</p> <p>For example, the software used at the family assistance center will assign each sample a number and the laboratory's accessioning program will assign a different number. Or, partner laboratories may assign their own number (and barcode) as they accession samples.</p>
Support shipping and receiving samples and data from partner laboratories.	<p>The LIMS should be able to:</p> <ul style="list-style-type: none"> <li>■ Build shipping manifests that contain samples or DNA extracts.</li> <li>■ Track the sample (or microtiter plate) as it moves among partner laboratories.</li> <li>■ Track when the DNA results and physical samples are returned.</li> </ul>
Segregate mass fatality data from regular casework.	The laboratory most likely will want to segregate the mass-fatality data from regular casework data so management reports and metrics are not merged and can be analyzed independently.

- **Identify conflicting results from different DNA technologies.** When multiple DNA technologies are used, the laboratory will need to review previously reported identifications to ensure that results from the new technologies are consistent. For example, a remains sample and a personal item may match with STRs but not with mitochondrial DNA (mtDNA).
- **Identify fortuitous matches.** Partial profiles resulting from sample degradation are a common occurrence in mass fatality incidents. A partial profile may match several reference samples fortuitously, particularly if the matching algorithm allows for the possibility of allelic dropout. The DNA analyst must review all of the candidate matches and choose an appropriate

course of action. The software should produce a work list that allows the DNA analyst to record free-text comments about each discrepancy.

If the laboratory chooses to outsource samples to partner laboratories, these additional quality control tools should be considered:

- **Data file validation.** The managing laboratory may want to validate the format and content of data files that are provided by partner laboratories. Fields that may be validated include sample names (to ensure the appropriate naming scheme was applied) and loci and allele values.
- **Blind-control verification.** One method of monitoring quality in partner laboratories is to



institute a blind-control program (see chapter 14, *Quality Control*). To partner laboratories, blind controls appear to be normal samples; however, their profiles have already been determined by the managing laboratory. The managing laboratory randomly places blind controls into the batches of samples (or microtiter plates) that are shipped to partner labs. The blind controls usually are renamed so that they are indistinguishable to the partner laboratories from normal samples. Then, the managing laboratory checks the data files that are produced by partner laboratories for blind controls and verifies them against the known DNA profiles.

### Matching and Statistics Software

There are two basic approaches to DNA matching: (1) direct matching, and (2) kinship matching. Direct matching compares two DNA profiles to determine whether they come from the same source (“individual”). Sophisticated direct matching algorithms consider allelic dropout for nuclear DNA and heteroplasmy for mtDNA. Kinship matching, on the other hand, uses DNA profiles to identify biological relationships among individuals. Kinship matching should consider both allelic dropout (nuclear DNA) and mutations (nuclear and mtDNA).

Exhibit 17 shows how mass fatality incident data may be searched.

One of the primary tools for making DNA identifications is “matching software.” Currently, the most widely used forensic DNA matching software in the United States is the FBI’s Combined DNA Index System (CODIS). However, an underlying design principle of CODIS is that matches are rare and independent events—and, in mass fatality incidents, matches are neither rare nor independent of one another. Therefore, a laboratory director should be aware of CODIS limitations in a mass fatality incident response.

CODIS is designed to rapidly search crime-scene DNA profiles against each other and against DNA profiles of known individuals. One assumption built into CODIS is that each profile will match only a tiny fraction (usually one or none) of the profiles in the database. In a criminal case, which CODIS is primarily designed to handle, the DNA profile obtained from a piece of evidence might not match any of the million-plus convicted-offender DNA profiles in the database, simply because that person has not previously been convicted of a crime that mandated collection of a DNA sample.

In a mass fatality incident, however, every human remain likely will match several samples, including other remains or personal items. Although CODIS can properly identify all of the matches in a mass fatality incident (through pairwise comparisons), it does not aggregate similar matches, and, therefore, is less useful in a situation where

#### Exhibit 17: Searching Mass Fatality Data

Versus (compared with)	Human Remains	Personal Items	Kinship Samples
Human remains	Use direct matching to identify multiple fragments of the same individual.	Identify human remains by direct matching to personal items.	Identify human remains by kinship matching.
Personal items	Identify human remains by direct matching to personal items.	Use direct matching to verify that multiple personal items submitted on behalf of a single victim have the same DNA profiles.	Use kinship matching to verify that the personal items belong to the victim.
Kinship samples	Identify human remains by kinship matching.	Use kinship matching to verify that the personal items belong to the victim.	Not usually performed.

the goal is to assemble all potential matches across time and space. That said, CODIS has a standard data file format that is used to report STR data, and this common “cmf” format was used in the WTC identification analyses.

Once a potential direct or kinship match is identified, the laboratory must determine its statistical significance using a likelihood ratio for kinship matching. To declare a match as an identification, the computed estimates must exceed threshold values that are predefined for direct and kinship matches. The identification thresholds are determined based on the number of victims, the biological relationships of the victims, and the nature of the incident. This was a major focus of the KADAP and is addressed in chapter 12.

Finally, the laboratory may elect to factor non-genetic data into the identification process. For example, human remains recovered from the WTC were catalogued based on their physical location within a two-dimensional grid superimposed on the disaster site. These data are useful when likelihood-ratio thresholds cannot be met due to incomplete DNA profiles.

## Sharing Information

Information technology (IT) should be considered both in the context of the identification work and as a tool to foster communication. Many laboratories have come to appreciate the value of a LIMS and sophisticated DNA matching/kinship software; however, IT also can be used to enhance interorganizational communication. The DNA laboratory will need IT solutions to exchange data with other government or vendor laboratories that are participating in the mass disaster response. For example, the ME most likely will have a system that tracks the status of all identification modalities, and the DNA laboratory’s system(s) should be fully integrated so data entry is not duplicated.

*Mass Fatality Incidents: A Guide for Human Forensic Identification* ([www.ojp.usdoj.gov/nij/pubs-sum/199758.htm](http://www.ojp.usdoj.gov/nij/pubs-sum/199758.htm)), recognizes some of these challenges:

The difficulty of [data management] is compounded when more than one laboratory is involved in providing DNA results. Participating

laboratories should affirm their mutual commitment, coordinate and track sample flow, and agree to use compatible software applications for data acquisition and interpretation.

When multiple organizations participate in a mass fatality response, several types of data may need to be exchanged, depending on the relationship between the organizations. External organizations and systems that typically require IT interfaces include:

- Government (partner) laboratories.
- Vendor laboratories.
- Sample collection agencies (e.g., first responders that collect and catalogue remains or collect personal items and kinship samples).
- CODIS.
- Mass-fatality-specific software programs (e.g., the Victim Identification Program (VIP), which is provided free of charge by the Disaster Mortuary Operational Response Teams).
- LIMS.

Exhibit 18 depicts business processes that will have to be integrated among laboratories if more than one laboratory is involved in the DNA identification analysis.

To successfully build interfaces across these business processes, the IT manager should consider:

- Physical connectivity between systems (e.g., Internet, private network, CD-ROM).
- Security requirements (e.g., encryption, firewalls, access controls).
- Data archiving.
- Communication mechanisms (e.g., file transfers, enterprise messaging, database interaction); most of this type of integration in mass fatality identification responses has been accomplished through file transfers.
- Data exchange formats (e.g., fields, field lengths, data types, relationships); data exchange formats are particularly challenging due to differences among some LIMS systems.
- Compatible barcodes that allow a barcode applied in one laboratory to be scanned and read in other laboratories.



### Exhibit 18: Integrating Processes in Multiple DNA Laboratories

Business Process	When Integration May Be Necessary
Sample accessioning	Multiple, independent computer systems are used to accession samples. All of the data must be accessible by the laboratory(ies) responsible for making identifications. The managing laboratory's LIMS may also need to communicate with DMORT's Victim Identification Program (VIP).
Sample analysis	One or more laboratories use the raw data produced by other laboratories to make its/their own allele calls. An example would be using an expert system to reanalyze severely degraded DNA.
Data interpretation	One or more laboratories review the allele calls made by other laboratories; participating laboratories produce CODIS-compatible import files for use by the managing laboratory.
Quality assurance	Multiple laboratories are participating in a single, unified quality assurance program.
Matching and statistics	Laboratories share their final results with one another.

### Infrastructure

Mass fatality incident responses are high-profile events that are scrutinized by the public, elected officials, and the press. In addition, the response can have an aggressive timetable for completing victim identifications, and DNA is often the primary means of identification. The systems that support the DNA identification effort should be considered "mission critical." System downtime should be minimized, and robust backup and restore procedures should be among the first processes implemented.

Volunteers or members of external organizations may participate in the laboratory's disaster response, which means that the IT manager may need to provide computers and other temporary services (e.g., printing and e-mail) to those entities. The IT manager most likely will need a security policy that restricts access to certain aspects of the data by unauthorized volunteers.

The demands placed on the IT infrastructure will last for the duration of the response. Because the DNA analysis process generates large amounts of data, the laboratory will need sufficient storage capacity to absorb the additional data produced

during the disaster response. Dedicated IT staff (in-house or outsourced) may need to be provided as well.

### Conclusion

Seasoned IT managers agree that building IT infrastructure (hardware and software) in the midst of a crisis is extremely difficult. Ideally, the laboratory will have in place the policies, processes, and network infrastructure to support a mass fatality incident response.

Prior to a mass fatality identification effort, the IT manager's strategic plan should:

- Identify agencies and systems that may require electronic interfaces and prepare a way to implement those interfaces.
- Create a procurement list of additional hardware and software to be purchased in the event of a mass fatality incident response, allowing for the fact that operating systems and hardware are continually changing.
- Ensure laboratory personnel know how to use software packages before a response is necessary.

- Ensure that the network infrastructure can be expanded quickly by adding new servers or desktop workstations.
- Develop a plan for adding additional, temporary IT staff.

The IT manager's strategic objectives during a mass fatality incident response should be to:

- Identify and control the design of interfaces to external systems.
- Ensure that adequate access controls are in place for external users.
- Provide reliable services.

Finally, if the laboratory cannot avoid writing custom software code during a mass fatality incident response, the IT manager should attempt to limit such software, because the introduction of new programming languages, platforms, etc., during this time increases the complexity of management, drives up costs, and can result in unexpected consequences affecting already functional programs.