Imagine standing trial for a crime you did not commit. Your innocence might be challenged solely by potentially biased “expert” testimony. This could be the case for many types of forensic evidence. The expert provides an uncompromising statement and offers no evaluation of the strength of the evidence or the threshold of evidential strength that led to the unqualified statement. Genetic-based evidence (DNA, RNA) has evolved well beyond unqualified statements to involve statistical assessment of the strength of the data. DNA-based methods are generally recognized as the gold standard of forensic science, a standard that may never be reached by some other disciplines in the field. The highly respected status of DNA evidence is partly based on the uniqueness of each individual and partly on the existence of databases to support the forensic applications. New DNA databases, such as the U.S. Y-STR database, continue to be created and evolve to support forensic practice. In other areas of forensic science, new databases are needed to support the advancement of operational forensic laboratories; however, progress in the non-DNA disciplines has been much slower. Large collections of information are required to move deficient evidential areas toward the data-driven forensics of the future. [1]

Much of the research at the National Center for Forensic Science (NCFS) is focused on building structured collections of information, known as databases, to assist in forensic casework today and transform current practices into the data-driven forensics of the future. This brochure highlights some of the databases developed at NCFS, the impetus for their development, social significance of the underlying forensic science and the resulting research. Many of these databases are accessed on a daily basis for forensic casework. New data collection needs are identified and incentivized based on predicted future trends in forensic science.

FIRE DEBRIS DATABASES

The importance of detecting arson crimes is reflected in the Florida Fires 2016 Annual Report. In 2016, the Florida Bureau of Fire & Arson Investigations conducted investigations on properties that had sustained collective losses estimated at $147 million due to damage from fires. Approximately $26 million of this estimated property loss was attributed to the crime of arson, which corresponded to 34% of the causes reported for all fires. [1]

According to a 2018 report from the Center of Fire Statistics, a total of 1.34 million fires were reported in the United States in 2016, which greatly exceeded the number of reported fires in any one of the 57 reporting countries throughout the world. [2] Fires in the U.S. resulted in 3,390 deaths and injured 62,085. On average, one in 10 fires took the lives of 69 firefighters and 62,085 inhabitants of the U.S. placed a call for fire service assistance in 2016.

On a global scale, there were 3.04 million fires reported in any one of the 57 reporting countries for an average of 3.50 million fires were reported in 57 countries. An average of 3.50 million fires were reported per year in the participating countries for years 2012 - 2016. According to a 2018 report from the Center of Fire Statistics, a total of 1.34 million fires were reported in the United States in 2016, which greatly exceeded the number of reported fires in any one of the 57 reporting countries throughout the world. [2] Fires in the U.S. resulted in 3,390 deaths and injured 62,085. On average, one in 10 fires took the lives of 69 firefighters and 62,085 inhabitants of the U.S. placed a call for fire service assistance in 2016.

On a global scale, there were 3.04 million fires reported in any one of the 57 reporting countries for an average of 3.50 million fires were reported in 57 countries. An average of 3.50 million fires were reported per year in the participating countries for years 2012 - 2016.

Prosecution and conviction for the crime of arson relies heavily on the forensic detection of ignitable liquid residue in fire debris. Equally important, accurate detection of ignitable liquid residue guards against erroneous arson convictions. The National Center for Forensic Science (NCFS) has created and maintains several databases that assist in the forensic analysis of fire debris and facilitate research efforts aimed at improving fire debris analysis.

The National Center for Forensic Science (NCFS) has created and maintains several databases that assist in the forensic analysis of fire debris and facilitate research efforts aimed at improving fire debris analysis.

The Ignitable Liquids Reference Collection (ILRC) was developed by TWGFEX and NCFS and became freely accessible on the internet in 2001. The ILRC database is a compilation of ignitable liquid analytical data and product information used by forensic analysts to conduct fire debris analyses. Records consist of unaltered, evaporated and biologically degraded examples of ignitable liquids. Ignitable liquid results from a March 1999 national survey of fire and explosion debris analysts conducted by the Technical Working Group for Fire and Explosions (TWGFEX) and NCFS established the need to create a national database to help laboratory analysts identify traces of ignitable liquids in debris from fires.

The Ignitable Liquids Reference Collection (ILRC) was developed by TWGFEX and NCFS and became freely accessible on the internet in 2001. The ILRC database is a compilation of ignitable liquid analytical data and product information used by forensic analysts to conduct fire debris analyses. Records consist of unaltered, evaporated and biologically degraded examples of ignitable liquids. Ignitable liquid analysis for the purpose of database creation is performed in accordance with standard forensic testing methods.

In casework, the database aids fire debris analysts in the interpretation of samples to determine a possible cause of a fire and enables the analyst to compare hundreds of possible ignitable liquids to their casework sample, including those altered by evaporation and biologically induced change. The ILRC is also an extensive training and educational resource for fire debris analysts.

In research, this large collection of data characterizes the commercially available ignitable liquids and offers a means to develop computer algorithms for predicting their presence in a casework sample. Importantly, the error rates for computer based (machine learning) methods can easily be determined.

International Database of Ignitable Liquids

Many laboratories performing fire debris analysis around the world have used the ILRC database; however, this U.S.-centric database did not always contain representative ignitable liquids from other regions of the world. The European Network of Forensic Science Institutes’ Fire and Explosions Investigation Working Group (ENFSI FEIWG), tasked itself with developing a database similar to the ILRC for European laboratories. The ENFSI FEIWG is comprised of members from 22 European nations.

The ENFSI FEIWG relied on NCFS to play a leading role in developing the new database as a resource for the international fire debris community. Participation of FEIWG members involved in creating the database was supported by the Prevention of and Fight against Crime Program of the European Union (EU). The resulting database, the International Database of Ignitable Liquids (IDIL), is a compilation of analytical data and product information from mainly EU forensic laboratories. Since its initial development, the IDIL has expanded to include samples from Australia and New Zealand. The database is housed on UCF servers and maintained by NCFS personnel as a service to forensic analysts around the world.

The IDIL facilitates cross-border collaboration among European forensic laboratories and international investigations in general. In addition, smaller laboratories without the means to support their own database now have access to hundreds of ignitable liquid records in the IDIL.

Restrictions on the international shipping of ignitable liquids can have a negative impact on research. The IDIL provides a source of research data and obviates the need for shipping samples internationally.
SUBSTRATE DATABASE

Following the successful launch and application of the ILRC, TWGFEX members requested a new database to assist in casework. The analysts needed a database containing information about the chemicals produced in a fire. During a fire, household furnishings and building materials (substrates) decompose under intense heat to produce chemicals, some of which may also be found in ignitable liquids. The fire-induced formation of chemicals complicates interpretation of the data. In response to analyst needs, the Substrate Database was developed by NCFS in collaboration with TWGFEX. The purpose of the database was to assist analysts in distinguishing fire-induced chemicals from trace residues of commercial ignitable liquids that may have been used to start a fire.

The Substrate Database is a compilation of analytical data and product information about the materials that commonly decompose to form fire debris. Database records consist of substrate materials that are unburned and burned to various extents. NCFS prepares and analyzes samples for the database in accordance with the standard testing methods used in forensic laboratories.

In casework, the database aids analysts in data interpretation by providing examples of the portion of a casework chromatographic profile that is contributed by the burning structure. In some cases, the entire chromatographic profile may be attributed to chemicals produced by the burning structure. The database allows the analyst to more easily determine if ignitable liquid residue is present in the debris.

Research performed at NCFS has electronically combined the records from the ILRC and Substrate databases to build models representing fire debris with knowledge of the sample’s composition. These digital models are used to train advanced machine learning algorithms to detect the presence of ignitable liquid residue in casework fire debris samples.

FIRE DEBRIS DATABASE

In order to validate fire debris data interpretation methods, data samples must be representative of casework samples and known to contain or not contain ignitable liquid residue. This is true for validating expert opinions or machine learning methods. The Fire Debris database was developed by NCFS to provide the data required to validate fire debris interpretation methods. This database represents a new resource for the fire debris analysis community, which allows the determination of error rates for this complicated area of forensic science.

The Fire Debris database contains chromatographic data and information from fire debris samples that are truly representative of samples that would be encountered in casework. In addition, the ground truth (presence or absence of ignitable liquid residue) is known for each sample. Records consist of examples containing exclusively burned substrate materials as well as mixtures of burned substrates and ignitable liquid residues. Samples in the Fire Debris database are composed of ignitable liquid and substrate contributions that are represented in the ILRC and Substrate databases. NCFS prepares and analyzes samples in accordance with standard forensic test methods.

The Fire Debris database serves as an educational resource for analysts and exposes them to samples with a similar level of difficulty as those they will encounter in casework. Analysts may also utilize the data in training for proficiency testing and validation of data interpretation methods.

In research, the data from the Fire Debris database provides a validation set for machine learning algorithms designed to identify the presence of ignitable liquid residue in fire debris samples.

SMOKELESS POWDERS DATABASE

The primary use of smokeless powders is the manufacturing and reloading of ammunition. Smokeless powders may also be used to construct improvised explosive devices (IED). In 1998, the National Research Council’s committee on Smokeless and Black Powder published a report on studies addressing the feasibility of adding tracer elements to smokeless and black powders. [1] Their recommendation was not to implement tracer elements, but to develop a national database that would assist investigators in identifying the manufacturer and product line of the powders. The Smokeless Powders Database was developed by NCFS in collaboration with TWGFEX in response to the recommendation for a national database of smokeless powders. The database enables explosives analysts to identify whether smokeless powders were utilized in a bombing incident. Once identification of a smokeless powder is established, the analyst can compare their casework sample to nearly 1,000 smokeless powders in the database to determine a potential product and manufacturer.

In the United States there were 647 explosion incidents reported in 2017, of which 335 were bombings. [2]


SEXUAL LUBRICANT DATABASE

The number of sexual assault cases with usable or no DNA evidence is on the increase. Results from analysis of 7,000 sexual assault kits (SAK) taken from National Institute of Justice funded projects reveal that DNA profiles of sufficient quality for searching against the combined DNA index system (CODIS) were obtained in only 2 in 5 cases. In the absence of quality DNA evidence, lubricants often used in sexual assaults offer an alternative form of trace evidence to assist in the investigation and prosecution of sexual assault cases. However, a database was required to fully realize the investigative and prosecutorial power of the evidence.

The Sexual Lubricant Database is a compilation of reference lubricants, characterization data, instrumental data and classification models that assist forensic scientists in conducting lubricant analysis in sexual assault cases. This database includes various types of lubricants that can be used to facilitate a sexual assault. Database use in casework is validated by the International Forensic Lubricant Working Group. This database is intended to be a reference and resource for the forensic lubricant community to provide a methodology of analyzing and classifying unknown lubricants collected in sexual assault kits, especially when quality DNA evidence is not present. Multiple analytical instruments and methods are involved in preparing database samples, which provides multiple analysis options for forensic laboratories using information in the database.

Records in the database also provide a resource for NCFS faculty and students to develop new lubricant classification schemes, and to assess the importance of various lubricant chemical constituents as class and individual characteristics for commercially available lubricants.

U.S. Y-STR DATABASE

The use of DNA in forensic science for the identification of individuals started with the work of Alex Jefferies in the mid-1980s. DNA methods have become the gold standard of forensic trace evidence. However, in some cases such as sexual assaults the evidence may include DNA contributions for two or more individuals. DNA mixtures are difficult to interpret unless something can be done to simplify the evidence. In male-on-female sexual assaults involving a male perpetrator, simplification may involve limiting the evidence. In female-on-male sexual assaults the evidence may include DNA contributions for two or more individuals. DNA mixtures are difficult to interpret unless something can be done to simplify the evidence. In male-on-female sexual assaults involving a male perpetrator, simplification may involve limiting the evidence.

The U.S. Y-STR Database, a searchable listing of 11- to 23-locus Y-STR haplotypes located on the web at www.ystrdatabase.org, has been developed by combining data from NCFS with databases from government, commercial and academic resources throughout the United States. The database provides tools for laboratories to obtain Y-STR haplotype frequencies needed to calculate matching or paternity probabilities with confidence intervals. Other features include the ability to simultaneously upload multiple haplotypes for searches directly from Genotyper® and GeneMapper® text files, the ability to include or exclude sampled populations and a report-style printout of the results.

ORLANDO DATA PORTAL

The City of Orlando curates an open data portal, which provides access to a database of Orlando Police Department (OPD) calls for service (CFS). The data portal is intended to increase transparency and facilitate economic development. [1] The CFS database contains records for calls for service of multiple types, as well as the geolocation (latitude and longitude) from which the call originated. Knowledge of where and when different CFS originated allows the prediction of where calls will come from in the future, which in turn allows OPD to deploy its resources in the most effective way to reduce response time and crime in the city.

The data contained in the Calls for Service database can be combined with associated data (population density, median house price, distance to the nearest major thoroughfare, etc.) and used to predict the probability of calls originating from specific areas based on the geographic pattern, economic status and other associated data. The map below shows the actual probability of domestic disturbance calls in greater Orlando for service in 2017 (right) and the predicted probability of domestic disturbance calls for service in 2017 using the previous five years probabilities, population density, median house price, and the distance to the nearest on-ramp for I-4, SR-408, and SR-417. Accuracy of the prediction is seen in the similar graph patterns with a small error between the predicted and actual probabilities of 0.01153.

As the population of U.S. cities grows, it is critical to develop smart infrastructures to support their growth. Smart cities enhanced by smart policing are possible today and research at NCFS is leading the way.

1. https://data.cityoforlando.net
Theft of credit card numbers by skimmers at the gas station pumps is an increasing problem in the U.S., resulting in a significant amount of theft each year. Approximately 2,000 skimmers have been found at pumps in Florida since 2015, with over 1,000 found in 2018 alone. [1]

A single skimmer has the potential to result in $1 million in fraudulent credit card charges. [2]

Identifying and removing individual skimmers is an important law enforcement activity. Tracking skimmer locations and characteristics in a database opens the possibility of connecting associated cases and identifying the individuals profiting from skimmers. In partnership with the U.S. Secret Service, state and local law enforcement, NCFIS is helping to build a Skimmer Database with the potential to attack the problem on a national scale.

This database will provide forensic practitioners with data on the skimmers, as well as when and where credit card information was taken. The data may also allow the prediction of future occurrences of skimmers to enable proactive policing in smarter communities of the future.
