Forensic Technology
CENTER OF EXCELLENCE
A program of the National Institute of Justice

IN-BRIEF SERIES (Part 2 of 3)

Digital Transformation of Cold Case Reviews: Digitizing Case Files

Introduction to Report Series

Advances in forensic science capabilities represent opportunities and challenges for allied professionals involved in cold case violent crime investigations. Modern technologies can uncover important insights that may not have been previously possible; however, law enforcement and other associated agencies can struggle to efficiently leverage the large amounts of information associated with a cold case. Factors such as a lack of case file organization, decentralization of documents, and the time-intensive process of manually searching paper records for relevant details make it difficult to identify cases that may benefit from the application of new techniques and technologies (i.e., new compared to what was available at the time of the initial investigation or advances made to what was available at the time of the initial investigation).

The cold case review process is a collaborative effort during which professionals representing a variety of disciplinary backgrounds review all relevant case details to develop a strategy that will advance the investigation. The members of this multidisciplinary team (MDT), which may consist of law enforcement investigators, forensic science service providers (FSSPs), prosecutors, victim/family advocates, medicolegal death investigators, and sexual assault nurse examiners (SANEs), share common challenges within the case review process.

> ***"There is no detective who doesn't recognize the need and value of digitizing cold case files."***
>
> *—Tom McAndrew, Pennsylvania State Police, Criminal Investigations Unit, retired; Lehigh County (Pennsylvania) District Attorney's Office, Homicide Task Force*

Objectives

- Communicate the benefits and realities of digitizing cold case files (with the help of case studies).
- Outline strategies that MDTs may employ to digitize their cold case data.
- Present considerations for MDTs implementing digitization technology.

Tracking down and extracting relevant information to create an informed forensic or investigative strategy is both time and resource intensive. MDTs may look to technology to digitize paper cold case records, thus enabling multiple agencies on a secured network to search, share, and access files. The digitization of case files may support the future implementation of text analytics tools, which are enabled by artificial intelligence (AI) and can quickly identify key details, relationships, and patterns within or between case file records. Applying AI to the cold case review process can help MDTs become more agile and efficient in collaborating and developing valuable forensic and investigative strategies, but this digital transformation requires technical, time, and resource investments. This in-brief is the second of a three-part series that highlights the potential value, approaches, and considerations for digital transformation of cold case files and the case review process. Although created for cold case MDTs with an emphasis on the role of law enforcement and FSSP involvement, this series is valuable for all forensic science service providers involved in an MDT and criminal justice decision-makers. All three in-briefs focus on sexual assault and violent crime cold cases and the possible benefits of identifying evidence for additional or advanced DNA testing, but converting paper records into searchable electronic records may also lead to the identification and (re)testing of other types of physical evidence to advance other types of

cases.[1] The information presented within these in-briefs can be applied to and benefit all cases regardless of the passage of time. **Specifically, this in-brief focuses on the processes for digitizing case files and the potential impact.**

Key Takeaways

- Digitizing information contained in cold case files—specifically, the text data contained in forms, narratives, and other documents—can streamline the case review process in many ways, enabling an MDT of allied professionals to store case data securely, search and access data easily within a single case and across different cases, and share information with collaborators.
- Fully digitizing cold case data involves scanning documents and converting text to "machine-readable" text that can be copied or searched. Tools that can convert text use optical character recognition (OCR) software technology are currently limited to converting typed text, not handwritten notes.
- MDTs can use manual transcription, off-the-shelf scanners and software, and contract services to digitize cold case data. Each approach has benefits, limitations, and investment requirements. This in-brief provides case study examples of organizations that have successfully digitized cold case files.
- Successful digitization efforts require front-end planning such as compiling relevant case files, organizing the paper files into an easy-to-reference system, and considering data storage options for the electronic case files.
- The process of digitizing case files often requires a resource investment and may include steps such as compiling case data from disparate sources and organizing digitized data outputs. Organizations may seek external resources for procuring the appropriate equipment, labor hours, funding, and other support.
- Digitizing cold case data is a necessary step to use emerging text analytics tools, which use AI technology to provide insights across large volumes of text-based data. These tools facilitate the understanding of key information, identify forensic testing opportunities, and recognize linkages across cases. Although current adoption of these tools for cold case applications is limited, digitizing data can pave the way for future implementation. More information about the value of these tools can be found in the third in-brief of this three-part series, "***Digital Transformation in Cold Cases: The Application of Text Analytics***."

Context

A significant number of investigations that have grown cold might be resolved with improved access to digitized data and analytics.

Resolving cold cases, or cases where investigative leads have been exhausted, is a challenge for law enforcement agencies, the criminal justice system, and the general public.[2] A 2008 survey conducted by the RAND Corporation revealed that one in five cold cases is resolved, and only one in 100 cases results in arrest leading to conviction.[3] As of 2020, there were an estimated 250,000 unresolved homicide cases within the United States, with lower clearance rates resulting in over 100,000 of these cases accumulating during the past 2 decades alone.[4] Furthermore, only 32.9% of rape cases were cleared by arrest or exceptional means by the investigating agency as of 2019.[5, 6a] Several factors may lead law enforcement agencies to initiate a case review process, including new forensic techniques and technologies, new case-related information, systematic reexamination of cold cases, and societal and external pressures.

Cold case review processes are often led by law enforcement investigators but should involve a collaborative working effort between MDT members. Certain scenarios may lend themselves to the inclusion of professionals such as a SANE or a medicolegal death investigator. For example, including representatives from organizations such as the National Center for Missing and Unidentified Children (NCMEC) or the National Missing and Unidentified Persons System (NamUs) in the MDT may be advantageous when reviewing cold cases involving unidentified victims or missing persons. In this case review process, the MDT often reviews large amounts of information gathered during the initial investigation. More information about the case review process can be found in the first in-brief of this series, "***Digital Transformation of Cold Cases: Prevalence, Challenges, and Benefits of Just Resolutions***."

For this in-brief series, the FTCoE has defined the following terms:

- **Cold Case:** Cases where investigative leads have been exhausted with specific focus on cases related to violent crimes, including homicide, sexually motivated homicide, and sexual assault.
- **Resolution:** An outcome resulting from a full investigation in which a case is cleared by arrest (i.e., an individual has been arrested/charged with the crime or the case is turned over to prosecution) or closed as a result of a circumstance outside of the investigating agency's control that prevents an arrest, charge, or take prosecutorial action against the individual (e.g., individual is deceased).[7]

Many cold case files are kept in paper formats, limiting their searchability, shareability, and security.

During an original investigation, agencies collect a large amount of case data, which can be in various forms: typed and handwritten notes, information on microfiche and microfilm (e.g., copies of documents or photographs), photographs and sketches, reports from SANEs or medicolegal death investigators, and other records. Most agencies revisiting cold cases follow a stepwise process to gather important information and decide whether to invest their resources into case resolution. In many law enforcement agencies, cold case files often remain in hardcopy form, leading to inefficiencies in the cold case review process. Investigators spend a significant amount of time reviewing paper case files, which can be

[a] The FBI's Uniform Crime Reporting Program defines rape as "penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim."[6] The reported static includes attempts or assaults to commit rape but excludes statutory rape and incest.[6] Clearance via exceptional means occurs when elements beyond a law enforcement's control prevent the agency from arresting and formally charging the officer (e.g., death of an offender, lack of cooperation from victims).[7]

hundreds of pages, to identify relevant case information. There may be a high volume of paper case files, making them difficult to share, and they may become misplaced or damaged without backup copies.

What "Digitization" Means for Cold Case Reviews

Digitization is a process by which an organization converts hardcopy files into a digital format. This allows investigators to organize, view, and analyze case files from a computer. There are varying levels to digitizing hardcopy data, but digitization has two key aspects (***Exhibit 1***).

Agencies may capture case data as scanned images as a means of digitizing; such images are not amenable to computer-based searching and analysis. **This in-brief looks specifically into tools that can help convert images into text.**

Why Digitizing Cold Case Files Provides Value for Agencies

Digitization may serve as an important step in streamlining processes for cold case reviews. Agencies that digitize their case data can benefit from the following:

- **Easy access and data sharing.** Investigators can view data in parallel and share information easily.
- **Assembly of centralized, searchable files**. With information organized by case number, investigators can search for particular keywords within one or more cases, which may provide a new perspective.
- **Enhanced security**. Digitizing guards against tampering, loss, or misplacement. Paper records may degrade with time, but digitized records will not.
- **Enhanced collaborative capabilities.** Digitized files may aid engagement with MDT member organizations.
- **Reduction in case file storage needs.** Some agencies may elect to convert and store files primarily in a digital format. Digitizing can free up file cabinet space and reduce dependence on paper file systems.
- **Future adoption of advanced text analytics.** Digitization converts case data into a format that can enable advanced searching. For example, investigators may be able to search for details like names, types of evidence, or types of crime, which may facilitate the investigation of leads more quickly. It also enables high-level text and file analyses so that investigators can quickly identify the number of cold cases and subcategories, like homicides that include elements of sexual assault. These techniques are discussed in the third in-brief of this series, "***Digital Transformation in Cold Case Reviews: the Application of Text Analytics***."

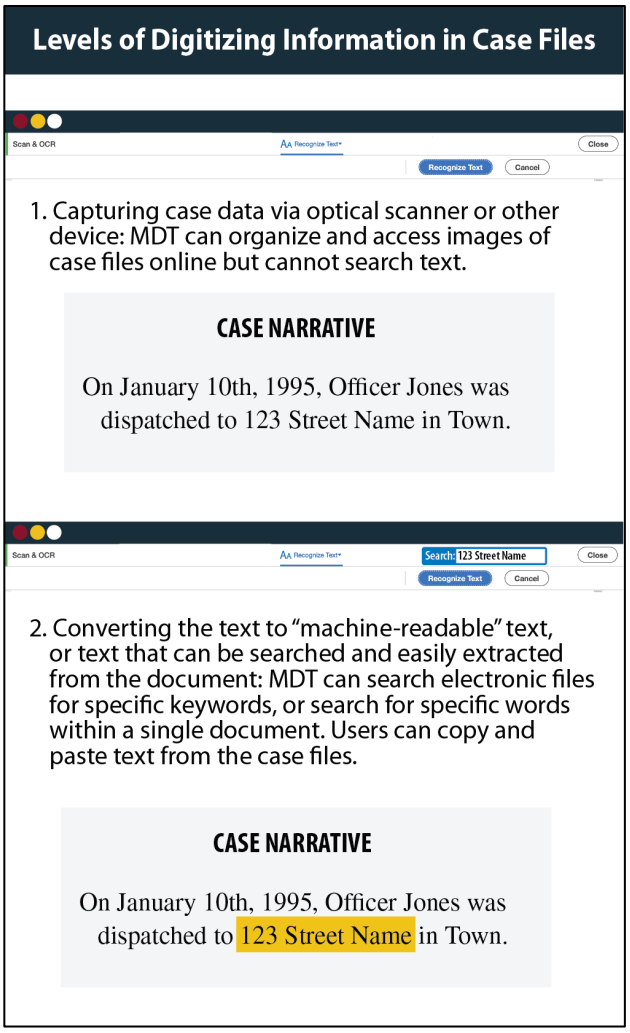

Exhibit 1. Converting to machine-readable text offers more searching capabilities than scanning files as images or PDFs.

Ten Steps MDTs Should Take Toward Digitizing Cold Case Files

Digitizing cold case data may require a committed investment by the agencies leading cold case reviews. Tool implementation is important for enabling successful digitization, as is up-front planning. Agencies should review their current workflow and cases to take the following next steps before investing resources:

1. **Understand and communicate the goals for digitizing cold case records.** Timely and just resolution of cold cases is a shared goal across the criminal justice community, and digitization may serve as a step to help achieve this goal.
2. **Determine the status of previous digitizing efforts.** Agencies should consider previous efforts to digitize case files and current policies for storing and accessing these data. To identify a clear path forward, agencies should understand the current state of their case files, their workflows related to current case management, and previous digitizing efforts. ***Exhibit 2*** provides examples of questions MDTs should consider to understand the following:
   a. The volume of cases that must be digitized and the time and potential cost requirements
   b. The pre-digitization steps needed to gather and organize the files
   c. How the digitized cold case data will be used

Cold Case Status	Workflow Status	Digital Status
▪ How many cold case files does the agency or unit have? ▪ Are these documents in good condition, with minimal damage? ▪ Where are these case files located, and how are they organized? ▪ What formats are these case files in (e.g., typed text, handwritten, microfiche or microfilm)?	▪ What are the agency's policies and procedures for digital data storage and sharing with MDT collaborators? ▪ What storage options does the agency offer (e.g., network, cloud-based storage)?	▪ Have files been scanned previously, and if so, into what digital formats, and where are they stored? ▪ What scanning equipment is currently available to the agency? ▪ Does the agency and MDT value the searchability of the data or just the security and aggregation into one easy-to-access place?

Exhibit 2. Questions an MDT should ask to inform a digitization strategy for cold case data.

3. **Invest time in aggregating case file data in one place.** Front-end data aggregation, regardless of the approach to digitize, is often the most resource-intensive step of the digitization process. This step may require significant engagement with other members of the MDT and across the criminal justice community.
4. **Consider the options for digitizing cold case files.** Agencies may pursue multiple approaches to digitize case data. Some approaches may use currently available scanning resources (and thus have no hardware cost) but require a labor investment. Many of these approaches leverage OCR, a technology used to convert printed, written, or typed characters into a digital format and into recognizable letters, numbers, and characters.[8] This allows the content to be "read" as text data that can be copied and searched. The next section covers four approaches to capturing and converting data that organizations may take:
   a. Manual transcription
   b. OCR-enabled software
   c. Integrated scanner systems
   d. Outsourcing to a Criminal Justice Information Service (CJIS)-compliant scanning company

5. **Invest time to check the quality of OCR text conversion**. Although OCR technology is robust, document quality or other factors may impact the accuracy of the text conversion. This is a significant challenge for agencies electing to replace their original records with digital ones. In the implementation phase, MDTs should cross-check OCR text conversions against the original file to ensure the software is accurately capturing information.
6. **Identify available resources and explore avenues for obtaining equipment, labor, or funding for this effort.** Digitization approaches often require a combination of labor and equipment, which may be challenging for budget-constrained agencies where cold case investigators are assigned multiple duties. MDTs may consider equipment sharing, external staff, and funding opportunities to help offset these costs.
   a. **Equipment.** Consider partner agencies, laboratories, and others that may have access to high-quality, integrated scanners. Other potential avenues include libraries, local universities, and archival associations. MDTs should ensure that public resources do not store memory and enable security of the data captured.
   b. **Staff or Volunteers.** Consider how external labor support may reduce clerical duties for investigators and increase their time to investigate cold cases. The following may provide access to "vetted" labor:
      i. Vidocq Society, a members-only group consisting of volunteer investigators and forensic experts serving as cold case consultants
      ii. Interns from local universities
      iii. Retired officers or military personnel
      iv. Community helpers
   c. **Resource Centers**. Federally or privately funded organizations may support law enforcement, FSSPs, and other collaborators by providing technical assistance or equipment.
   d. **Funding.** Consider how external funding (e.g., federal and state grants) can assist. Funding programs from the Bureau of Justice Assistance (BJA) and other federal, state, or local agencies may be able to support digitization efforts.
7. **Identify the pricing model that fits the needs defined by the MDT.** There is a tradeoff of either time or money saved between in-house and outsourcing options. Therefore, it is important for an MDT to consider how agency resources are invested—whether in-house labor hours and equipment costs outweigh the price of outsourcing. There are additional values to consider, such as convenience of investing in a scanner or digitization system that can be used in the future. Although physical scanners have incrementally improved over the past few years, software products to process, convert, and manage scanned data have evolved rapidly.

NCMEC realized the value of organizing digital case files as they digitized roughly 8,000 paper files over the course of a year. They learned that:

- Organizing is key to digitizing case files successfully. Creating a document organization system is easier to do with paper files first, rather than digitizing and organizing retroactively.
- Within each case file, organizations should delineate a chronological order for documents to be scanned so that MDTs can easily track down information. This information may also allow for quick input into external databases, such as NamUS.
- Labeling digitized documents with important information, such as the associated case file number and document type, helped rectify instances where case files were digitally "misplaced" on the network.
- Creating a step-by-step best practices manual with detailed instructions can help the entire organization digitize documents and save files consistently.

8. **Consider organization of case files prior to scanning.** Although digitizing case files offer benefits such as ease of searching and accessing files and reducing storage space, the benefits of digitizing may only be realized when these files are organized in a consistent and easy-to-navigate format. This requires up-front planning and agency-wide collaboration.
9. **Consider where to store electronic case files.** To emulate the security, organization, and data hygiene practices of hardcopy case files, these data must be properly stored and managed in accordance with the policies and procedures of the agency. Storage options include the following:
    a. **Local network storage.** For this approach, files can be organized in folders on network drives and are accessible to computer users via virtual private network (VPN). MDT members can search for keywords using file explorer and easily track down well-organized files, but they may not be able to see who has accessed or moved data.
    b. **Integration into a records management system (RMS).** MDTs could decide to add cold cases to an agency's RMS, a database used for storing, retrieving, and retaining records.[9] In many law enforcement agencies, current cases are managed in a case management module of an RMS. These systems offer security and audit trail capabilities and may enable searchability features that can help MDTs find relevant information quickly. Importing data into the RMS may require a significant time investment beyond digitizing. More information about the functionality of RMS can be found in the Law Enforcement Information Technology Standards Council's **Standard Functional Specifications for Law Enforcement Records Management Systems, Version II**.
    c. **Storage via custom document management system.** Many document scanning outsourcing firms offer a document management option, where they can build a database for agencies to store and access files. These companies can help agencies store data in a location that aligns with their policies and procedures and access data with a user-friendly interface. These systems may be expensive to create and maintain and may necessitate ongoing maintenance contracts with the document scanning company.
    d. **Cloud-based storage.** As MDTs gather more data associated with investigations, such as body camera footage and other forms of digital evidence, many agencies are moving to cloud-based storage to keep up with the data demands. These environments are secure, and some providers, like Microsoft and Amazon, offer CJIS-compliant cloud storage options. Agencies that have already implemented cloud-based services for data storage may look into storing case data in their cloud-based storage, where they can use tools within product suites to further search and analyze the machine-readable content (e.g., **Amazon Comprehend**, a text analytics tool, for data stored in the Amazon Web Services cloud).
10. **Consult the expertise of peer organizations that have undergone digitization initiatives**. Some MDT members from other organizations across the United States have partially or fully completed the cold case digitization process and can provide guidance on how to navigate the process from start to finish. Later, this in-brief provides three case studies that summarize the digitization efforts undertaken by the NCMC, New Jersey State Police, and the Los Angeles Police Department.

NCMEC, for example, provides resources to help law enforcement, FSSPs, and other collaborators review cold case files and develop investigative strategies; however, NCMEC requires these files to be digital prior to this collaboration. Additionally, they offer to purchase scanners with digitizing capabilities if the agency they are working with does not have the capability within their funds. Having digitized their own paper files, the organization is able to support law enforcement agencies and FSSPs in need of digitizing their files.

Approaches for Digitizing Cold Cases

MDTs may consider one or more approaches to digitizing their case data. Approaches that convert typed text rely on OCR technology in the form of widely available OCR software programs. Traditional OCR, however, is generally limited to printed text. No current tools effectively convert handwritten notes. Although some technology companies have had limited success in converting handwriting in designated "blocks," like a suspect's name and address, these methods cannot account for the variability between handwritten text in semi-structured or unstructured documents. Handwritten notes must be scanned as images, which are not searchable, or transcribed manually.

A variety of documents may be located in case files, such as typed documents, handwritten notes, and documents stored on microfilm and microfiche. These documents could contain any of the following:

- **Structured data:** Information is neatly displayed in tables or in pre-defined fields of a form. Formatting, number, and layout are completely static between documents. Examples include questionnaires, names, dates, and numbers.
- **Unstructured data:** Information is not organized in a pre-defined manner. Examples include free-written notes or text files, PDFs, presentations, microfiche, and media that was not filmed or photographed with a mobile device.
- **Semi-Structured data:** A mixture of structured and unstructured data. Examples include (1) email, in which the text body has no structure but the name, email address, or time stamp are structured, and (2) digital photographs from a cellphone, which may be geo-tagged and include time stamps. FSSP reports with structured fields and free-form written or typed notes are also examples.[10]

Exhibit 3 provides an overview of major approaches and technologies that organizations may employ to digitize their cold case files. Please note that the following product and service examples are meant to highlight, not promote or endorse, any vendor. These examples do not represent a comprehensive summary of available products and services.

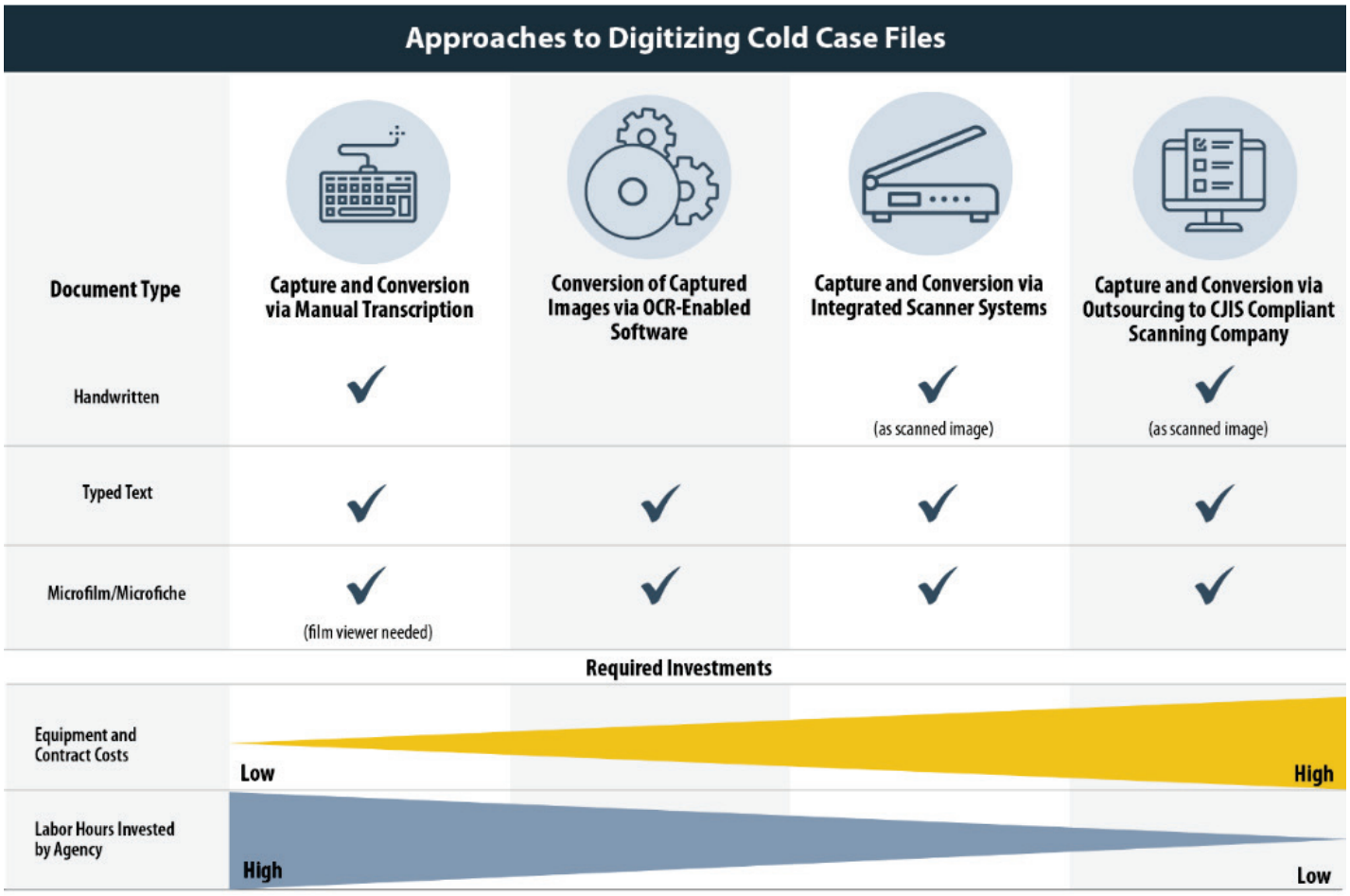

Exhibit 3. MDTs may consider one or more approaches to digitizing case files, depending on the type of paper files they possess and their budget for equipment and labor costs.

Capture and Conversion via Manual Transcription

MDTs may consider physically transcribing case documents via Microsoft Word or Excel. This is the best option for handwritten notes, which cannot be converted via OCR technology. Agencies can use this for typed documents or forms.

Application and Value

MDTs may consider this option if…

- Most of their paper files or microfiche or microfilm content contain handwritten information.
- They are looking to make case file information machine-readable.
- They would like to do additional analysis while digitizing cases (e.g., create case summaries).
- They have labor hours available to support manual transcription (whether in-house or through vetted volunteers).

INVESTMENT

Although this solution does not require purchasing any additional scanners or software packages for digitizing, it will likely require hiring an additional staff member, intern, or volunteer because this is a labor-intensive process.

Benefits

- Training individuals to transcribe is a low investment.
- Agencies do not need to invest in scanning hardware or software packages.
- Transcribers can review important information in cases as they manually transcribe and may use this as an opportunity to electronically code or tag files to make them more easily searchable.

Limitations

- Manual transcription has a long turnaround time and depends on case file volume and personnel.
- This technique can be liable to data entry errors made by those transcribing content.
- MDTs with documents stored on microfilm or microfiche need to invest in a reader to view the information.

Procurement Examples and Resources

- Most transcription services focus on audio/video to text tasks, and the FTCoE found no services that offer handwriting transcription services that are CJIS-compliant.
- The National Archives Citizen Archivist Dashboard provides **this resource page** with guidance for transcription and tagging photographs for identification purposes.
- Microfiche and microfilm readers are machines that allow an individual to view the image from microfiche or microfilm on a screen. Today's models of these readers have both computer and internet connectivity, allowing digital records to be viewed, printed, and digitized.

Conversion of Captured Images via OCR-Enabled Software

OCR-enabled software products can convert scanned images of typed text to machine-readable text. These products, which are available as apps or software product suites, can be used directly after scanning paper documents or can convert already scanned images.

Application and Value

MDTs may consider this option if…

- Most of their documents contain typed text.
- They have already scanned their typed documents as high-quality images, and want to make this text machine-readable.
- They have access to high-quality scanners and enough labor hours to convert these images.

INVESTMENT

Many of these products are priced as yearly subscriptions and may cost around $50–$200 per year/license. Agencies must also plan for time investments for scanning and converting each file using the software.

Benefits

- The MDT may be able to use simple, inexpensive scanners that they already own.
- Typed text case information already stored as images can be converted to machine-readable text.
- By separating scanning and conversion steps, agencies can streamline the scanning step by using multiple scanners.

Limitations

- The image must have been scanned by a scanner with sufficient dots per inch (DPI) or a resolution to effectively capture text and other elements like photographs or sketches.
- Implementation may require set-up to add conversion software into the digitization workflow.

Procurement Examples and Resources

Users can choose from a variety of software products, including the following:

- Products that are part of a greater suite of document capture and processing tools such as **IBM Datacap**, **Microsoft Azure**, **Amazon Textract**, and **ABBYY FineReader**
- Lower-cost software applications that primarily perform an OCR task such as **Iris Readiris**, **Kofax OmniPage Ultimate**, and **Meta Enterprises, LLC SimpleOCR**.
- Embedded capabilities as part of a greater general product suite such as **Microsoft OneNote** and **Adobe Acrobat** are capable of converting PDFs or images to text and may already be used in agencies and FSSPs.

NCMEC used Adobe Acrobat to convert scanned text-based files into machine-readable format when they digitized 8,000 paper files. More information can be found in the ***Examples of Cold Case Digitization Efforts*** below.

Capture and Conversion via Integrated Scanner Systems

Integrated scanner systems consist of a high-resolution scanner with built-in OCR software. The system has a built-in workflow that enables users to scan documents and covert to machine-readable text in one step.

Application and Value

MDTs may consider this option if…

- Most of their documents are typed text, or if they would like to scan documents with handwritten text as images.
- They need to digitize documents in microfiche or microfilm format.
- They have money available for equipment and labor hours.

INVESTMENT

There is a wide price range for OCR-enabled scanners, starting as low as $300 for those with a 50-page feed capability. The price of these scanners typically increases with higher document feed sizes (e.g., a 200-page feed capability printer costs roughly $3,500), there is an additional investment in personnel required to organize, sort, and scan the files.

Benefits

- These scanners can concurrently scan and convert text using OCR, streamlining the digitizing process.
- Automatic document feeding options can accelerate the rate of digitization.

Limitations

- No known systems can perform both microfiche or microfilm and typed text scanning.
- Integrated scanners may be more expensive than conventional scanners, which may make justifying procurement to decision-makers difficult.

Procurement Examples and Resources

- Scanning companies like **Fujitsu**, **Brother**, **Epson**, and **Canon** have recognized the need for OCR and have created off-the-shelf integrated scanning systems with OCR capabilities or OCR software packages. These OCR-enabled scanners typically have an automatic document feed feature, which reduces the amount of time required by personnel to feed papers by hand.
- There are OCR scanners for both typed text, such as the previously mentioned companies, and microfiche or microfilm, such as the **ScanPro** scanner series.

The New Jersey State Police leveraged a Fujitsu FI-900 J scanner to digitize over 200 of their cold case files in-house. More information can be found in the ***Examples of Cold Case Digitization Efforts*** below.

Capture and Conversion via Outsourcing to a CJIS-Compliant Scanning Company

An MDT may consider outsourcing efforts to collect, prepare, and scan documents with OCR to create a repository of digitized data. Many firms regularly digitize records for criminal justice community members and implement a strict workflow protocol to meet CJIS compliance standards.

To begin the process, the vendor typically requests a pilot set of information to run a proof-of-concept trial. This ensures the client and company are aligned with the desired outcomes for the scanning process.

Before scanning, the documents are prepared by removing paperclips and staples, and the pages are inspected for quality issues. Once the initial inspection is complete, the papers are scanned into a digital version with an OCR filter and checked for quality. Finally, the files are indexed, or named and organized according to the agency's goals. Digitized files are returned to the agency in either a "document dump" or are delivered in a database system created by the vendor. In addition, there are CJIS-compliant companies that offer microfilm or microfiche conversions.

Application and Value

MDTs may consider this option if…

- Their case files include typed text, microfiche, or handwritten text (scanned images, though some may offer transcription services).
- They have not yet started the scanning process.
- They have a large backlog of paper case files.
- They have the budget but lack the resources to digitize in-house.

INVESTMENT

Service pricing may include a per-page cost model, starting roughly around $0.10/page. This cost could go up to $0.20/page if OCR scanning is added. Additional fees such as service and set-up are not included.

Benefits

- The MDT can enable digitization of case files even during busy periods with limited staff availability.
- MDTs can reduce some (but not all) time and labor investments for digitizing.
- Some companies can scan and convert case files at the agency's site.

Limitations

- This conversion service is expensive compared to other approaches.
- MDTs may need to either make copies of data or risk potential loss of data during transport to vendor facility.
- Agencies may not trust their sensitive data in the hands of contractors.

Procurement Examples and Resources

- There are CJIS-compliant companies that offer microfilm or microfiche conversions (e.g., **BMI Imaging**).

The Los Angeles County Police Department leveraged the Federal Bureau of Investigation's (FBI's) scanning capabilities when digitizing their "murder books." More information can be found in the ***Examples of Cold Case Digitization Efforts*** below.

Examples of Cold Case Digitization Efforts

With the success of their own in-house digitization experiences, NCMEC now provides resources to smaller agencies considering digitizing their case files.

NCMEC assists the criminal justice community with long-term investigations, including cold case investigations involving child victims, by providing case consultations and other resources.[11] In 2017, NCMEC began the process of digitizing approximately 8,000 paper case files. The organization pursued digitization of their files as part of an effort to transfer their large volumes of paperwork to an electronically secured network and reduce physical paperwork during an office move. To begin this process, the team first started with a small group of case files as a test case to determine a consistent approach. This enabled NCMEC to optimize their document scanning protocol—during this process, the team completed the following steps:

1. Prepared, reviewed, and organized case files.
2. Digitized documents using a leased high-speed scanner that were saved in PDF format.
3. Digitized and saved non-document data (media such as photos or videos) in the appropriate format (e.g., JPEG, MP4). Most of this was performed in-house, but NCMEC used the U.S. Secret Service Forensic Science Laboratory to digitize films and VHS tapes.
4. Opened each scanned document with typed text in Adobe Acrobat, which uses OCR to convert text to machine-readable format.

The process of aggregating, organizing, and scanning case files took approximately a year to complete. Although many individuals from NCMEC were involved in this process, they also used consultants, interns, and NCMEC management for additional help. Throughout this process, NCMEC had an internal spot-checking system to ensure that digitization was being completed correctly.

"We always advocate law enforcement records to be digitized. It increases sharing, collaboration, and preservation of case files"

—Carol Schweitzer, Supervisor of Forensic Services Unit, NCMEC

Digitizing these files allowed NCMEC to bring each of the case files up to current investigation standards. Once the files were digitized, NCMEC was able to search across case materials using internal data mining tools. For example, a suspect's name could be searched using this internal tool and mentions or searches across any other cases within the NCMEC system where this suspect was named would populate.

During this digitization process, NCMEC benefited from organization-wide buy-in with motivated personnel. NCMEC now uses this in-house digitization experience to help smaller agencies digitize their files, as many larger agencies already have the required resources. NCMEC provides additional help to agencies, including case consultations, introductions, or other resources like purchasing a high-speed scanner for digitization.

NCMEC also sees the digitization of these case files as a step toward future adoption of AI-enabled tools for understanding connections between cases. Please see the third in-brief of this series, "***Digital Transformation of Cold Case Reviews: The Application of Text Analytics***," for more information.

Front-end investment in digitizing cold cases offered the New Jersey State Police the ability to streamline, organize, and dive into high-priority cases.

Detective Sergeant Joe Itri and the New Jersey State Police used funding from BJA's National Sexual Assault Kit Initiative (SAKI) to digitize over 200 of their cold case files with the help of an integrated scanner system. Their Cold Case Unit took the following steps:

1. Prepared the case files by locating all relevant information, sorting the files, and removing paperclips and staples.
2. Fed papers into the integrated Fujitsu FI-900 J scanning system (borrowed from their crime laboratory), which scanned the files as PDFs and saved into a standalone, off-network computer.
3. Transferred files into a computer connected to their agency network.
4. Input these files into their case management system.

> **"Scanning is just one part of the process–this just means that we have digitized our information. To gather meaning, we need to sort and analyze the data."**
>
> *— Joe Itri, Detective Sergeant First Class, New Jersey State Police*

Although aggregating and preparing the files required a significant time investment, the process of scanning the documents took around 50 hours in total. The New Jersey State Police did not require any additional hires for labor inputs; rather, the officers digitized the files as they had the time or worked overtime. Using a scanner with a high automatic feed capability, the agency was able to digitize roughly 1,000 pages every 15 minutes. This scanner was enabled with OCR capabilities and converted all written text documents (forms, typed notes, news articles) into text that could be searched for names and other keywords. The unit also had manually transcribed handwritten ledgers that were used to manage the evidence book into a Microsoft Word document, which enabled them to search for evidence associated with a certain case number.

Digitizing these files served as just one step of a workflow that helped the agency sift through cases, capture relevant summary data, and analyze the aggregated data. With these files, the New Jersey State Police was able to:

- **Efficiently develop an understanding of status across many cases.** During the digitizing process, investigators reviewed files to identify and revisit cold cases with a sexual assault nexus. In this process, the unit created a Microsoft Excel spreadsheet summarizing each case and the case status to help investigators quickly understand the case at a high level. Examples of captured data points are victims' names and ages, date and location of the crime, case numbers and laboratory report numbers, and evidence resubmissions. Investigators could use targeted searching to recall data to populate the spreadsheet.
- **Enhance capacity for deep analysis of cases.** Front-end investment for aggregating, digitizing, and extracting relevant case information reduces demands on the investigator. With centralized, comprehensive, and easy-to-search data, investigators can spend their limited time reviewing case details.
- **Recall case details in a timely manner.** In circumstances where external factors (like media attention) have brought up old cases, investigators can quickly query their database to find relevant information.

In addition to the BJA SAKI grant funding, the New Jersey State Police had agency-wide buy-in, with motivated personnel to spend time and resources on their efforts to digitize cold cases. This buy-in and motivation were invaluable to their ability to move their efforts forward.

Digitizing homicide files was a key step in enabling the Los Angeles Police Department to use their standardized homicide investigation material for advanced data analysis.

The Los Angeles Police Department (LAPD), one of the country's largest police agencies, developed a practice to systematically document homicide investigations in the 1980s, which has led to a large body of standardized case information. For each homicide investigation, the department organizes case file data into standardized binders the LAPD refers to as "murder books." Each of these binders is divided into 26 sections and can contain anywhere from 800 to 1,000 pages of information. The sections are then further organized by tabs. For example, Tab 1 includes the chronological record written by the investigation team. Other sections contain the crime report, victim and suspect information, and other written entries.

The comprehensiveness of case data documentation positioned the LAPD as an ideal candidate for research projects related to homicide solvability. A necessary first step to use these data was to digitize the paper files. The LAPD began these efforts by gathering unresolved and resolved homicide cases, spanning from 1990 to 2010, from the Operations–South Bureau region (an area that typically has the highest number of homicide cases).[12]

The LAPD partnered with the FBI, consulting firm Justice & Security Strategies, Inc., and the University of California Los Angeles to digitize and eventually analyze the case data. Funding from the National Institute of Justice (NIJ; 2018-75-CX-0003), the BJA (2018-WY-BX-0002), and the Ahmanson Foundation enabled the LAPD to create a strong digital foundation of case files, which has led to important follow-on work (see the third in-brief of this series, "***Digital Transformation of Cold Case Reviews: The Application of Text Analytics***," for more information on this work). The LAPD used the following process to digitize their files (see ***Exhibit 4*** for a visualization of this process):

1. LAPD gathered, boxed, and shipped the binders (murder books) to FBI Headquarters.
2. The paper files were digitized and then returned to the binders.
3. The digitized files were burned on DVDs, and both the DVDs and binders were shipped to the LAPD.
4. Once digitized, consulting firm Justice & Security Strategies reviewed the files on each DVD. This included ensuring the files were complete and all data had been digitized.
5. Each case was tagged by variables such as whether the case was resolved or unresolved, the motive, and the relationship of the suspect(s) to the victim(s), improving the keyword search capacity of the database.
6. Once reviewed and tagged, the files were uploaded to the Homicide Library System, allowing investigators to access the case files from a computer.

The LAPD and FBI constructed a 7,300 square foot Homicide Library, which opened on October 2, 2019. This library, built with funding from the Ahmanson Foundation, houses both the physical murder books and the network where the digitized case files reside. This enables the city-wide assembly, retrieval, and digitization of LAPD's homicide cases.

> ***"No matter what, being able to digitize something is a necessary first step."***
>
> *—Dr. Craig Uchida, President of Justice & Security Strategies, Inc.*

Currently, an estimated 4,304 murder books from the Operations–South Bureau region have been digitized and are housed in the Homicide Library. Eventually, over 15,000 murder books from all 21 divisions of the LAPD will be digitized and housed in this central location. This allows the LAPD to:

- **Track current and cold case files properly**. As the library evolves, investigators will be allowed to "check out" the physical murder books or have the files emailed (if digitized).
- **Search the database.** When all the case files are digitized and the database is complete, investigators will be able to search many aspects of a homicide case file, including license plates, names, and locations.[13]
- **Develop advanced investigation tools**. Digitization facilitates research projects that ultimately aim to improve data analysis during homicide investigations. See the third in-brief of this series "***Digital Transformation of Cold Case Reviews: Application of Text Analytics***" for more information on the ongoing efforts regarding this database.
- **Provide services to the victims' loved ones.** The Homicide Library provides a place where a victims' loved ones can schedule a time to speak with an investigator on the status of the case.[14]

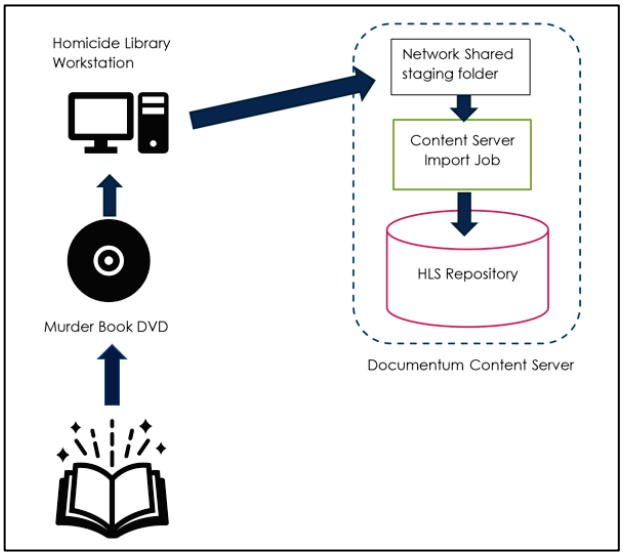

Exhibit 4. Digitization workflow of the LAPD Homicide Library (graciously provided by Justice & Security Strategies and supported by NIJ award 2018-75-CX-0003 and BJA award 2018-WY-BX-0002).

Conclusion

Digitally transforming paper case files to searchable, shareable, and secured electronic data is technically feasible for virtually any organization. Although the digitization process requires some front-end strategizing for data aggregation and organization, MDTs can leverage many approaches to achieve this goal. In many cases, MDTs may be able to lean on products or services already available. This process will not happen overnight—digitization may take from a few months to a year,[b] depending on the tools used and the resources available.

Digitizing represents only one step of digital transformation: digitizing is necessary to eventually implement advanced technologies, such as text analytics and other tools that may streamline the cold case review process. Although these tools and approaches are nascent as of this in-brief's publication, they may play a key role to MDTs in the future.

For more information on how MDTs may prepare for future technology implementation, refer to the third in-brief of this series, "***Digital Transformation of Cold Case Reviews: The Application of Text Analytics***."

[b] Based on FTCoE-conducted interviews with organizations that have digitized their cold case data.

Appendix: Glossary

Digitization: The process of digitizing; the conversion of analogue data (e.g., images, video, text) into digital form.[15]

Dots Per Inch (DPI)/Resolution: Refers to the sharpness and clarity or an image, typically indicating the DPI. When a file is digitized, the page is split into a grid of boxes. Each box is represented by a 0 or 1, depending on the information contained within. The result is known as a bit map. The denser the bit map (i.e., the more boxes a page is split into), the higher the resolution. For example, a 300 DPI printer is capable of printing 300 distinct dots in a line 1 inch long; this means 90,000 dots per square inch are printed.[16]

Machine-Readable text: Data that can be read and processed by a computer.[17]

Optical Character Recognition (OCR): The technology used to convert printed, written, or typed characters into a digital format and into recognizable letters, numbers, and characters. This allows text to be read by a computer so that characters can be edited and searched. Having an OCR system is required to edit text read by an optical scanner.

Optical Scanner: A device that can read text or illustrations on a page and transform the information into a form that the computer can understand.[16]

Semi-Structured Data: Data that have a loose organizational framework; a mixture of structured and unstructured data. Examples includes email, in which the text body has no structure but the name, email address, or time stamp are structured, and digital photos from a phone, which may be geo-tagged and include date and time stamps. Laboratory, medical examiner, or coroners reports with structured fields but free-written or typed notes are also included.[18, 10]

Structured Data: Data that are highly organized in a pre-defined format that can be easily understood by machine learning. These data are most often categorized as quantitative data and are the data with which we are most used to working. Examples include names, addresses, structured forms, or data existing in pre-defined formats.[19, 20]

Text-Based Analytics: The process of using AI to translate large amounts of unstructured text in documents into quantitative data. This is useful for uncovering insights, trends, and patterns.[21, 22]

Unstructured Data: Data that are difficult to deconstruct because they have no pre-defined model or structure; therefore, these data cannot be organized into relational databases. They are most often categorized as qualitative data. Examples include text, video files, data not contained in a structured form, social media posts, or email.[23, 24]

References

1. Forensic Technology Center of Excellence. Beyond DNA – Sexual Assault Investigations Prefix *Beyond DNA – Sexual Assault Investigations.* In-Brief Report Series. Vol. 2022, Washington, DC.: National Institute of Justice, 2019. https://forensiccoe.org/beyond-dna-reports-sexual-assault-reform/.
2. Heurich, Charles. "Cold cases: Resources for agencies, resolution for families." National Institute of Justice. July 14. 2008. Last modified February 28, 2022, https://nij.ojp.gov/topics/articles/cold-cases-resources-agencies-resolution-families.
3. Davis, R. C., Jensen, C., and Kitchens, K. E. Cold-case investigations: An analysis of current practices and factors associated with successful outcomes Prefix *Cold-case investigations: An analysis of current practices and factors associated with successful outcomes.* The RAND Corporation, the National Institute of Justice, 2011. https://www.rand.org/pubs/technical_reports/TR948.html.
4. Martin, Eric, Schwarting, Dawn Elizabeth, and Chase, Ruby J. Serial killer connections through cold cases Prefix *Serial killer connections through cold cases.* National Institute of Justice, 2020. https://nij.ojp.gov/topics/articles/serial-killer-connections-through-cold-cases#note3.
5. Federal Bureau of Investigation. "2019 crime in the United States." Accessed February 28, 2022. https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/clearances.
6. U.S. Department of Justice, Federal Bureau of Investigation. "Rape." Accessed May 16, 2022. Federal Bureau of Investigation, https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/rape.
7. U.S. Department of Justice, Federal Bureau of Investigation. "FBI — Offenses Cleared." Accessed Mar 30, 2022. Federal Bureau of Investigation, https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/clearances.
8. TechTarget Search Content Management. "OCR (optical character recognition)." Accessed February 28, 2022. https://searchcontentmanagement.techtarget.com/definition/OCR-optical-character-recognition.
9. Bureau of Justice Assistance, and National Institute of Justice. Law enforcement records management systems (RMS) Prefix *Law enforcement records management systems (RMS).* Bureau of Justice Assistance, National Institute of Justice, n.d. https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/leitsc_law_enforcement_rms_systems.pdf.
10. Marr, Bernard. "What's the difference between structured, semi-structured and unstructured data?" Forbes. 2019, https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/.
11. National Center for Missing & Exploited Children. "What is the CyberTipline?" Accessed February 28, 2022. https://www.missingkids.org/HOME
12. LAPD. "Grand Opening of the LAPD & FBI Homicide Library;LAPD Murder Investigation Books NA19139sf." Accessed Mar 30, 2022. Los Angeles Police Foundation and the LAPD, https://www.lapdonline.org/newsroom/grand-opening-of-the-lapd-fbi-homicide-librarylapd-murder-investigation-books-na19139sf/.
13. Santa Cruz, Nicole. "Victims' families hope for answers in LAPD's homicide library." Los Angeles Times. November 19. 2013, https://www.latimes.com/local/la-me-c1-homicide-library-20131119-m-story.html.
14. Santa Cruz, Nicole. "Joint FBI and LAPD homicide library brings new hope for old cases." October 2. 2019, https://www.latimes.com/california/story/2019-10-02/lapd-and-fbi-to-announce-opening-of-homicide-library.
15. WhatIs.com. "Definition: Digitization." Accessed February 28, 2022. https://whatis.techtarget.com/definition/digitization
16. Webopedia. "Resolution." Last modified May 24, 2021. Accessed February 28, 2022. https://www.webopedia.com/definitions/resolution/.
17. Open Data Handbook. "Machine-readable data." Accessed February 28, 2022. https://opendatahandbook.org/glossary/en/terms/machine-readable/.
18. MonkeyLearn. "What is semi-structured data?" Accessed February 28, 2022. https://monkeylearn.com/blog/semi-structured-data/.

19. Smallcombe, Mark. "Structured vs unstructured data: 5 key differences." ingrate.io. January 3. 2022, https://www.xplenty.com/blog/structured-vs-unstructured-data-key-differences/.
20. Pickell, Devin. "Structured vs unstructured data – What's the difference?" G2. November 16. 2018, https://www.g2.com/articles/structured-vs-unstructured-data.
21. MonkeyLearn. "Text analytics basics: A beginner's guide." Accessed February 28, 2022. https://monkeylearn.com/blog/what-is-text-analytics/.
22. TIBCO. "What is text analytics?" Accessed February 28, 2022. https://www.tibco.com/reference-center/what-is-text-analytics.
23. NLP-progress. "Relationship extraction." Accessed February 28, 2022. https://nlpprogress.com/english/relationship_extraction.html.
24. Meel, Vidushi. "What is semi-supervised machine learning? A gentle introduction." viso.ai. n.d., https://viso.ai/deep-learning/semi-supervised-machine-learning-models/.

Published November 2022

More Information

FTCoE Contact

Jeri Ropero-Miller, PhD, F-ABFT
Director, FTCoE, RTI International
jerimiller@rti.org

NIJ Contact

Jennifer Love, PhD, D-ABFA
Physical Scientist
Office of Investigative and Forensic Sciences
Jennifer.love@usdoj.gov

Technical Contacts

Patricia Melton, PhD	Rebecca Shute, MS
RTI International	RTI International
pmelton@rti.org	rshute@rti.org

Disclaimer

The NIJ FTCoE, led by RTI International, is supported through a Cooperative Agreement from the NIJ (2016-MU-BX-K110), Office of Justice Programs, U.S. Department of Justice. Neither the U.S. Department of Justice nor any of its components are responsible for, or necessarily endorse, this in-brief. NIJ is the research, development, and evaluation agency of the U.S. Department of Justice. NIJ is dedicated to improving knowledge and understanding of crime and justice issues through science. NIJ provides objective and independent knowledge and tools to inform the decision-making of the criminal and juvenile justice communities to reduce crime and advance justice, particularly at the state and local levels. The NIJ Office of Investigative and Forensic Sciences (OIFS) is the federal government's lead agency for forensic science research and development. OIFS's mission is to improve the quality and practice of forensic science through innovative solutions that support research and development, testing and evaluation, technology, information exchange, and the development of training resources for the criminal justice community.

Public Domain Notice

Suggested Citation

Shute, Rebecca, Brailey Faris, Matthew Mecray, Mikalaa Martin, Ashley Rodriguez, Shannon Krauss, and Patricia Melton. "Digital Transformation in Cold Cases; Digitizing Cold Case Files." Forensic Technology Center of Excellence. U.S. Department of Justice, National Institute of Justice, Office of Investigative and Forensic Sciences, November 2022.