NAACCR Data Destruction Primer

Version 1.0

Editors:

David Chesnut, Information Management Services, Inc.

Bozena M. Morawski, Cancer Data Registry of Idaho

Castine Clerkin, North American Association of Central Cancer Registries

Steven Friedman, National Cancer Institute, National Institutes of Health

Susan Gershman, Massachusetts Cancer Registry

Selina Khatun, Nunavut Cancer Registry

Lauren Maniscalco, Louisiana Tumor Registry

Robert McLaughlin, Cancer Registry of Greater California

Recinda Sherman, North American Association of Central Cancer Registries

Qianru Wu, Nebraska Cancer Registry

Publication date [1] : November 12, 2021

[1] These guidelines should be reviewed and updated every 12 months.

Table of Contents

Introduction

On the surface, data destruction, also known as the deletion of data, would seem like a very straightforward topic and something easy to do. If you no longer want or need some piece of data, you can simply hit the delete key on your keyboard and it is gone. Easy, right? The data must be gone since I can no longer get to it! However, that could not be further from the truth. Data that have been saved on almost any electronic device, such as a server, laptop, desktop, mobile phone, or removable storage, are almost definitely NOT "gone" by simply pressing the delete key. This is because of the way the underlying technology of devices store data.

To make sure that data we want or need to delete are really gone, we need to understand a little bit about what is really going on when data are deleted and how we can pre-emptively take steps to make sure data we want gone are gone for good.

Scope

This document will attempt to provide a high-level understanding of why data destruction or media sanitization are important and some best practices that can be followed to ensure that sensitive data that must be destroyed are really destroyed.

This document will not attempt to go into all the details or things to consider about data destruction or media sanitization, since that has already been documented by other people. (See ***NIST SP 800-88*** on media sanitation for in-depth guidelines on media sanitation.)

This document also does not address rules and contractual obligations regarding record retention, which may be very registry- or contract-specific.

Finally, this document also does not explicitly address data destruction with respect to paper records, although some best practices for the handling and destruction of electronic media also apply to paper records (see **Media Destruction** below).

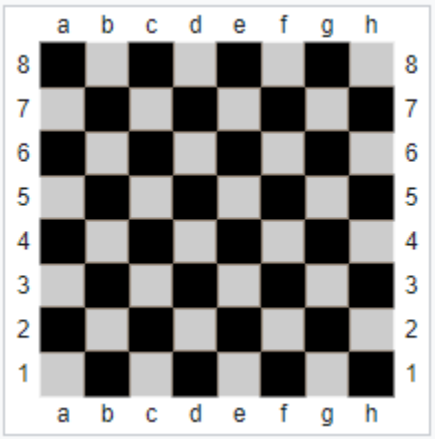
Definitions

Term	Definition
HDD	Hard Disk Drive, a storage device with mechanical parts that uses magnetic spinning “disks” to store data.
SSD	Solid State Disk, a storage device with no mechanical parts that contains nonvolatile flash memory to store data.
DoD 5220.22 M Standard	A standard method of data erasure which specifies a process for overwriting HDDs with a pattern of ones and zeros to ensure all data have been erased from the HDD.
NVM Express or NVMe	NVMe stands for Non-Volatile Memory Express and is a protocol designed to talk to SSD storage devices.
ATA Secure Erase	A low-level command that is built into some SSDs and HDDs that allows you to erase the drive and reset it to factory defaults. The preferred way to erase an SSD.
Shredding	The act of physically destroying a device or a paper copy of data. Typically performed by a certified company.
End of Life	The point at which a piece of hardware is considered outdated and no longer useful.
Aging Out	This refers to files in backup sets where data over time will naturally be removed from the backups when older backup sets are deleted because of retention policies.

Why Data Are Not Gone When We Delete Them

While we don't want to explore the inner workings of Hard Drives or Memory Cells, I do want to convey a general understanding of why data are not necessarily gone when we delete files from our computers or devices.

To explain this, we must first understand how files are stored on electronic media. In general, files are broken up into smaller pieces and then written in chunks to different places on the electronic media. In addition to the pieces being written to the media, an index that describes where all the pieces for each file are located is also written.



One way to think about this is using a chess board as an analogy for electronic storage media. We can think of each square on the chess board as a storage location for data. For our purposes, let's say that each square can store 10k of data. If we want to store a file that is 30k, it will take up 3 squares. Since each square on a chess board has a number starting from the lower left corner (a1) and going to the top right corner (h8), we can imagine that our file is stored at a1, b2, and c3. To keep track of our files, we also have a piece of paper where we write down all the locations of each file we store and the order of how all the pieces go together. Additionally, we also keep a list of the currently available squares that we can use to store any new files. If we keep doing this to store files, we will end up with a bunch of files stored on our chess board storage device and written down on our piece of paper so we can retrieve them when we want them. Now let's imagine we want to delete a file that is stored on c5, d8, and h2. To accomplish this, we can simply erase the file from our sheet of paper and then add those blocks to our available block list. Our paper will now tell us the c5, d8, and h2 are "free" to be used for storage again. However, the original data stored on those blocks is still there, it has not been deleted – we just "forgot" that it was there. This is exactly what electronic storage devices do when we delete files stored on them. They don't erase the file; they just erase the entry in what is called a File Allocation Table to "forget" where the file was stored.

While the above explanation is a bit simplistic, it illustrates how most storage devices work and why data are not necessarily gone when we press the delete button. The following sections will provide some things to think about, best practices, techniques, and some general information about available utilities.

Best Practices

This section will attempt to provide some "Best Practices" for data destruction and/or media sanitization. Some may seem obvious, while some may not.

Have a Plan and/or a Policy

This is the #1 best practice. To effectively delete data and sanitize media, the organization needs to have a plan **BEFORE** the data is stored – I cannot stress this enough. Trying to come up with a plan after the fact can lead to some unintended consequences, and, many times, a lot more work.

This plan typically involves having policies and procedures that detail what type of data can be stored, and where and how systems storing that data (laptops, desktops, servers, phones, and removable devices) will be handled when they reach their "end of life."

Many of the other Best Practices will expand on this "Have a Plan" section, and again, I cannot stress enough how important having a plan is.

The following are some of the items that should be considered for inclusion in a data storage policy and/or a media sanitization procedure:

- Decide on the reuse of storage devices. Will you allow devices such as server hard drives, laptop SSDs, or mobile phones to be wiped and then reallocated, or will they need to be destroyed?
- How will you "end of life" a piece of hardware that stored data? This typically comes into play with network or server-based storage. Will the drives from the device be wiped, or will they be shredded or otherwise destroyed?
- Where will really sensitive data be stored? Can there be a special area setup ready to receive data containing PII? See the "Categorize your Data" best practice for more on this.
- What will you do with backup media? See "Dealing with Backups" best practice for more.
- What will you do with received media?
- How will you transfer sensitive data, and do you need to consider the transfer mechanism in your data destruction plan?

While the above is not an exhaustive list of things to consider for a data handling plan, these points provide a good cross section of what should be considered.

Categorize your Data

Categorizing your data is the #2 best practice. If you do not know what kinds of data you have, you cannot effectively have a plan on how to protect them and how to delete them when they are no longer needed. I'm not suggesting that a full-blown data categorization solution is needed (there are ones available out there). However, having simple policies that consider if including Personally Identifiable Information (PII), such as SSN, name and address, or Protected Health Information (PHI) in data is a must. (See ***NIST SP 800-122*** on protecting the confidentiality of personally identifiable information.) Once the data are categorized, the policy can direct the user about where and how the data should be stored. For example, the policy can stipulate when certain data must be:

- stored encrypted

- only stored on a limited access network share
- only stored on machines with self-encrypting drives

Always storing sensitive data in specific locations or in specific ways allows those responsible for protecting and deleting them to have standard ways to manage data over their lifetime.

Data Backups

Sensitive data within backups can be challenging to delete. There are very few backup programs that allow a backup administrator to pick and choose files out of the backup set to destroy. Many times, it is an "all or nothing" scenario. This can mean that if you are required to delete "all instances" of a particular data file, you will need to delete all backups containing that file – which could be against your institute's policies, because you would not have any backups of the other files.

There are several ways to handle backups of sensitive data, and this is where planning comes in:

- Back the files up separately. This is not as easy as it sounds. Having separate backups is very inefficient to manage. This can be done, but your network administrators will not be happy with you if you ask for this.
- Store the files encrypted with the password stored in a safe place (i.e., in a password manager). Make sure only the encrypted version of the file is backed up. When the file needs to be used, unencrypt it into a directory that is never backed up.
- Exclude backups from any data destruction agreement or policy. This can work because backups typically "age out" over time, meaning that data no longer on the file system will not be in the saved backups after a period of time. This does require you to understand how long that retention period is and accept that. Most backups age out after 1 month or 1 year, but you should double check if this is the case with your organization. This also requires you to understand how backup media are handled. Are they encrypted? Are they destroyed when at their end of life? How long are they reused?

As you can see, even a little bit of planning, especially when it comes to backups, can prevent a huge headache down the road.

Use Self-Encrypting Drives / Full Disk Encryption

While this is not a paper specifically on data security, I would be remiss if I did not discuss the benefits of Full Disk Encryption (FDE) and Self-Encrypting Drives (SEDs) as they pertain to data destruction.

FDE is typically associated with software-based applications that encrypt data before they are placed on the storage device. For example, Microsoft's BitLocker can be configured on the professional versions of Windows to provide FDE. The BitLocker application is responsible for encrypting all the data before they are sent to the Hard Disk or SSD of the machine running the software. FDE can provide data security and secure erasure of the device once the device has hit its end of life by telling the application to "forget" the encryption key used for the encryption.

Self-Encrypting Drives (SEDs), on the other hand, do not require software to perform the encryption, but instead have hardware-based encryption engines that handle it transparently. SEDs are always good choices for storage devices even if they are not configured to provide data security.

SEDs, as the name implies, always encrypt data before they are stored on the device. However, SEDs without a password act just like a regular drive. Without a password, they do not prevent access (security) to the data stored on the drive, even though the data are physically stored encrypted. If you would like to use SEDs to secure data, I suggest using a key manager to set the password on the SED and handle passing it to the device automatically. There are multiple software vendors that provide this type of application. Microsoft BitLocker can do it on a PC-by-PC basis. However, for institutions, I would suggest a centralized key management application that can handle all the devices in one place so that those passwords don't get lost – if they are lost, you cannot get to your data.

I stated previously that SEDs are always a good idea when dealing with data destruction. The reason for this is even if you do not set a password on the device, the device stores all the data in an encrypted form. Because the data are encrypted, and all drive manufacturers provide a way to regenerate the key used for that encryption, we can easily wipe a drive cryptographically just by asking the drive to change its key! It is fast, efficient, and, most importantly, secure.

Use File Encryption

The final best practice I would like to discuss is file encryption. While encrypting a file is usually associated with protecting the file from being read by unauthorized people, it is also a great way to make sure that when data are deleted, they are truly gone. When a file is encrypted and the user "forgets" the password, the file has essentially been deleted because the data are unrecoverable. It doesn't matter where that file is – if the password (which should have been securely managed) is deleted, the file can be removed in the normal way, and it can be considered securely deleted. I touched on this earlier in the "Data Backups" best practice.

Destroying or Sanitizing Media

Whether or not you have followed the Best Practices listed above, at some point you will need to securely destroy or sanitize the media that held sensitive data. This section will provide some guidance for that process.

Media Destruction

The safest way to ensure that sensitive data cannot be recovered from any media that they have been stored on is to destroy the media. This can be as simple as running a CD/DVD or Floppy Disk (remember those?) through an appropriate physical shredder or taking a drill and a hammer to a hard drive (not recommended, but still effective and sometimes cathartic).

Typically, the shredding company you already use for secure paper shredding can also handle the shredding of all sorts of electronic media. Many will provide certification of media destruction or will even do it onsite so it can be witnessed. The main advantage of using a shredding company is that they know what they are doing and will usually recycle the shredded material afterwards.

Sanitization of Non-SEDs and Other Non-encrypted Devices

Disk drives and other non-encrypted devices must be sanitized utilizing software that writes random data overtop all the existing data multiple times. This can be a lengthy and time-consuming process. The larger the device is, the longer the process will take. The software must write the data multiple times (usually at least 3) to be compliant with many federal standards, most notably DoD 5220.22-M. This can take hours, if not days, to complete for some devices. Many times, it is much more cost-effective to use the shred option listed above and buy a new device rather than to try and repurpose it by wiping.

There are software options for securely wiping devices listed in the Tools & Resources section below.

Sanitization of SEDs and Devices with Full Disk Encryption

Sanitizing a Self-Encrypting Drive (SED) or a device with FDE is accomplished by changing the encryption key used on the drive. This is true whether the drive has a password on it to prevent access or not. For SEDs, there is software listed below in the Tools & Resources section that can be used to securely wipe them. For FDE devices, the software used to encrypt the drive can "forget" the key.

Sanitizing Mobile Devices

How a mobile device is "properly" sanitized depends on how the device was originally set up. If the device was set up with Full-Device Encryption from the start, sanitization is as simple as resetting the device and following the prompts. For devices that were encrypted, just change the encryption key on the data stored on it. If the device was not encrypted, there is no good way to re-initialize that device and know for certain that sensitive data is not still accessible somewhere on the device. However, as of the writing of this document, I believe that most mobile devices now default to being encrypted.

One item that should be considered when sanitizing mobile devices is if the device is backed up to "the cloud." This should be considered before allowing sensitive data to be stored on that device. However, if cloud backups were done, it will need to be addressed in some way.

The last item that can be used for sanitizing mobile devices is a mobile device management application. These are applications that are managed by the organization and create encrypted space on the device which can be wiped independently from the rest of the device. This is a suggested solution if your organization heavily relies on mobile devices to store and access sensitive data.

Items to Consider When Providing Data to an Outside Organization

While you cannot control what someone else does with data once they leave your organization, there are some things that you should ask for, or get assurances on, if you are providing data of a sensitive nature to outside organizations. This list is by no means meant to supersede intuitional policies or practices – it is just a list of things to consider.

- Does the organization have a Data Security Policy that includes how media and data will be handled after their useful lifetime?
- If there is a data destruction requirement, does that requirement cover backups? Are you okay with backups "aging out" of the other organization's system?
- How are the sensitive data files being transferred?
    - Are they encrypted?
    - Are they on media? If so, what happens to that media after the files are copied to permanent storage?
- How will you ensure verification of file deletion if it is required?

While this is not an exhaustive list, it does provide some key questions to think about.

Takeaways

Data destruction is not something that should only be considered when you need to do it, but something that requires forethought and a plan. When there is a plan, data can be easily and securely removed, and you can be sure that it was done correctly. Without a plan, data deletion can take some time, cause unintended consequences, and might not even be possible to do.

Tools & Resources

The following is a list of some tools that may be useful for the sanitization of drives, destruction of data, and further insights into other technologies. NOTE: This is not an endorsement of any of the products; they are only provided for reference.

Blancco Drive Erasure

This is the commercial version of DBAN (which is licensed for personal use only). Note that the enterprise version allows for erasure of many different drive technologies, so it provides a single solution for erasing "most" drives and even other media. https://www.blancco.com/products/drive-eraser/

Parted Magic

This is a disk management solution that also includes a secure erase component. It only works on SSDs and HDDs.

https://partedmagic.com/

Linux client software for SSDs

If you run a Linux based operating system the NVMe-CLI software will allow you to securely erase NVMe compatible SSDs directly from the Linux command line. This software is available on most Linux distributions. The following link provides an overview of the software.

Nvmexpress.org client software

Manual SSD sanitization

Most new SSD provide a sanitize option built into their onboard operating systems. This means that they support standard commands that make sanitizing the drive, either by erasure or changing of the cryptographic key (recommended), easier. However, many times this requires entering the host computer's BIOS to execute the commands. I do not recommend this for the normal user.

SSD utilities from manufacturer for older SSDs

Many manufacturers provide their own utilities to erase older SSDs that had proprietary commands for drive sanitization. The following are some manufacturer-specific utilities. You will need to determine the drive manufacturer and the drive model before checking their website for erasure tools.

- Intel Solid-State Toolbox
- Corsair SSD Toolbox
- SanDisk SSD Toolbox
- Samsung Magician Software
- OCZ Toolbox

Further Information on Self-Encrypting Drives

The following is an article from the Trusted Computing Group (TCG). They provide standards for Trusted Platform Modules (TPM), Trusted Network Communications (TNC), and Self-Encrypting Drives. This may help you make decisions of how to handle data management and destruction for your organization.

https://trustedcomputinggroup.org/resource/self-encrypting-drives-sed-overview/

Encrypted Hard Drives in Window

This article provides an overview of using encrypted drives on the Microsoft Windows operating system. It provides some useful information that you may want to consider for your organization.

https://docs.microsoft.com/en-us/windows/security/information-protection/encrypted-hard-drive

References

National Institute of Standards and Technology (NIST) Special Publication (SP) 800-88 Rev. 1 Guidelines for Media Sanitization

National Institute of Standards and Technology (NIST) Special Publication (SP) 800-122 Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)