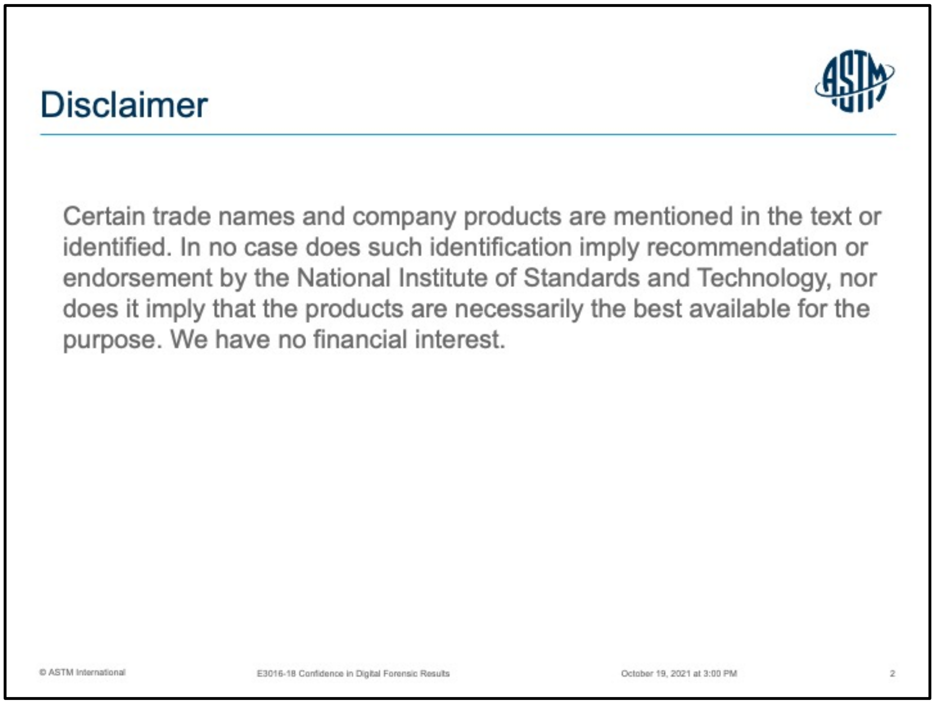
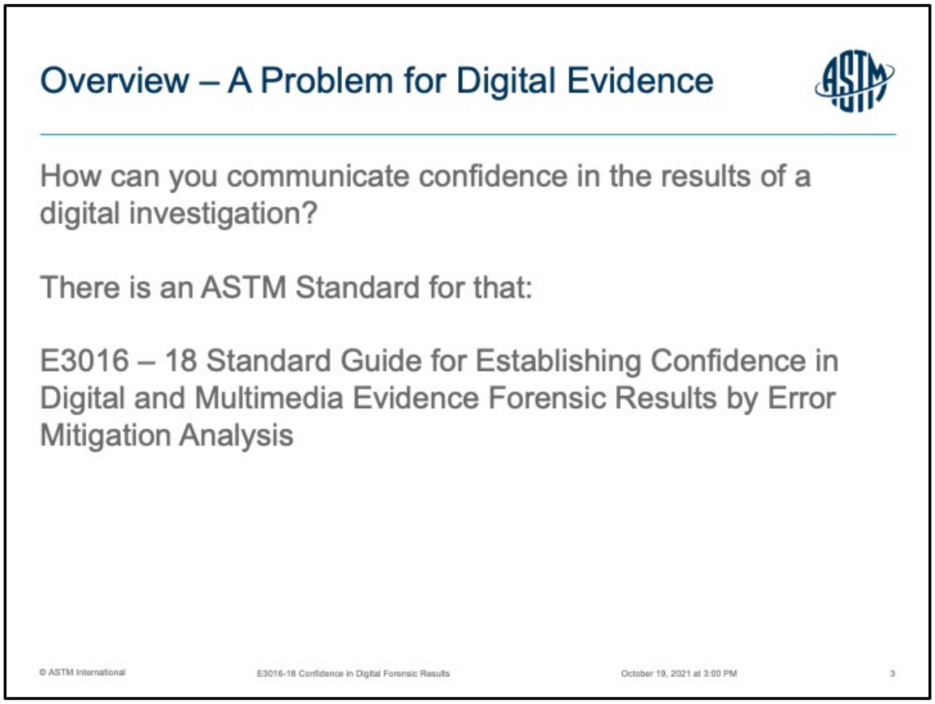
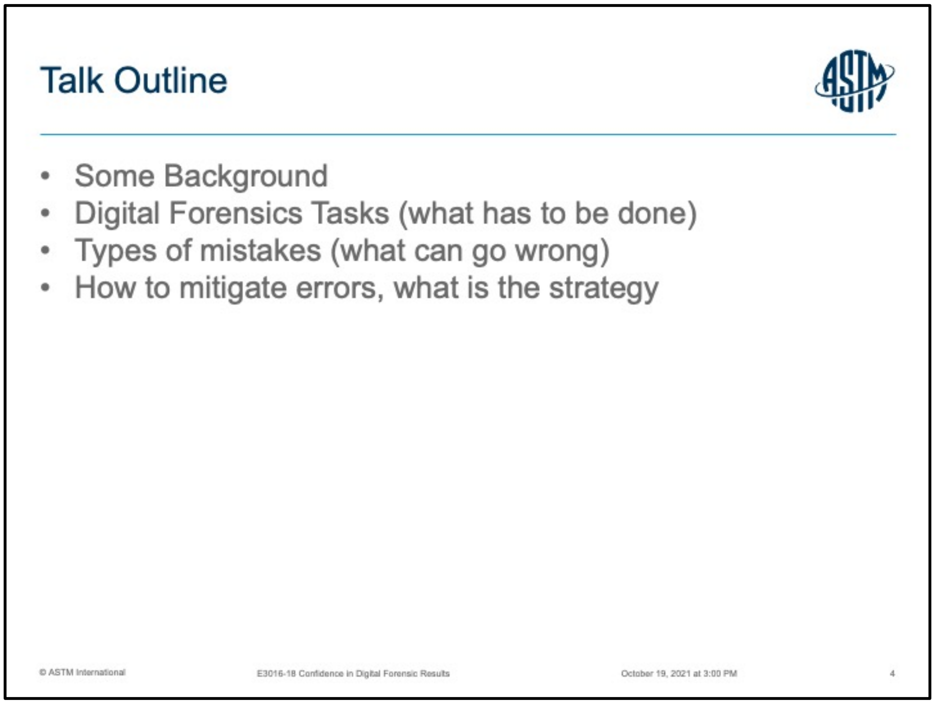

Good afternoon. Thank you for attending this talk.




I will try not to mention any specific products in my talk. If I do mention something I do not have any financial interest in these products.




This talk is based on the ASTM E3016-18 “Standard Guide for establishing confidence in digital and Multimedia Evidence Forensic Results by Error Mitigation Analysis.“ The document was originally written as a SWGDE guideline and then submitted to ASTM.




Here is an outline of today's talk.

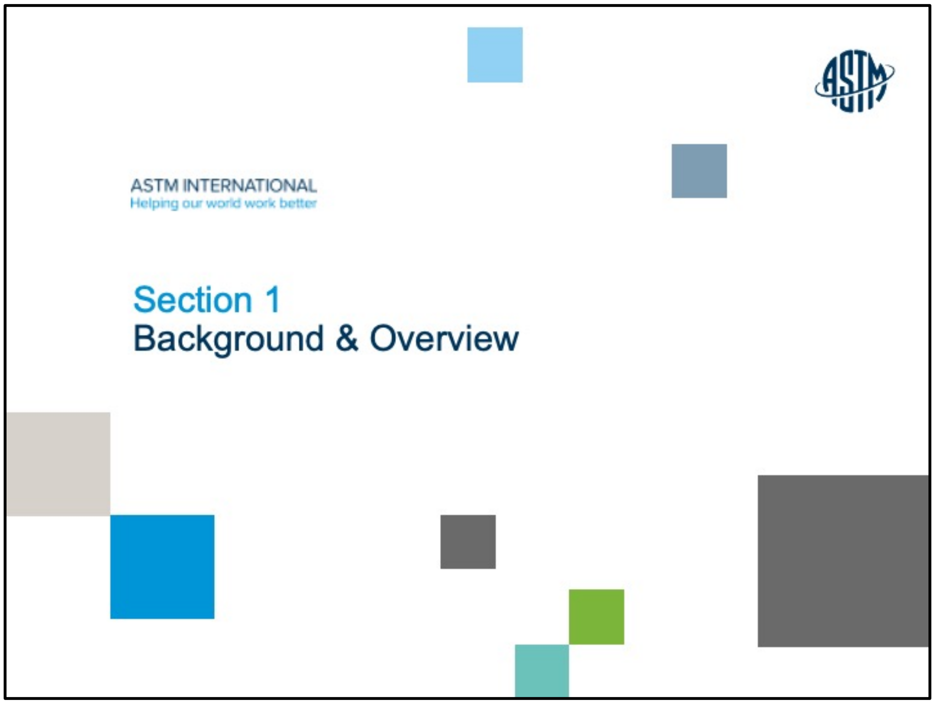
I'll talk about ways to characterize reliability of results

I'll talk about some tasks in other fields that focus on a single technique that can be described by an error rate, however digital forensics needs more than an error rate because so many different tasks make up a digital investigation. For example, DNA forensics may focus on a small set of questions like Does a sample from the crime scene match a sample from the suspect? Another important question for DNA is treatment of a sample that is a mixture. This is a current topic of research.

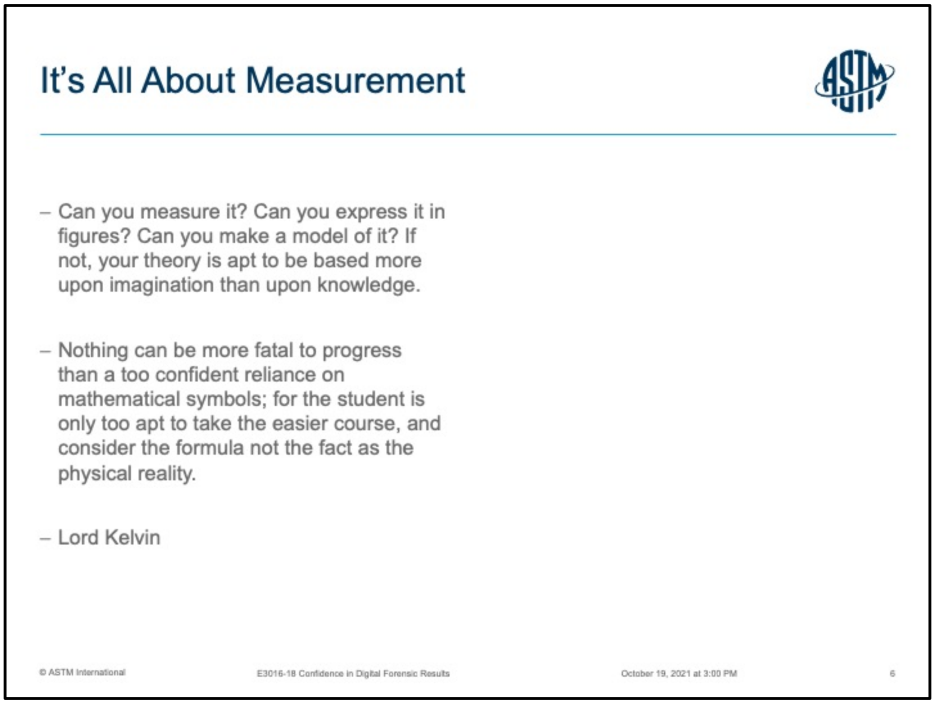
For digital, you might need to use several independent techniques, e.g., use a hash to identify a file of interest, a keyword search to locate a file about a topic of interest, recover a deleted file, etc. Of course, you can state an error rate for each one, but there can be quite a few independent tasks that makes in difficult to aggregate an error rate for the entire investigation.

Each digital tool is based on an algorithm designed to do a task that often can be characterized by an error rate. Sometimes these error rates for digital algorithms are so small as to be essentially zero. However, there is a hitch, the algorithm must be implemented in software and in the process systematic errors can be introduced.

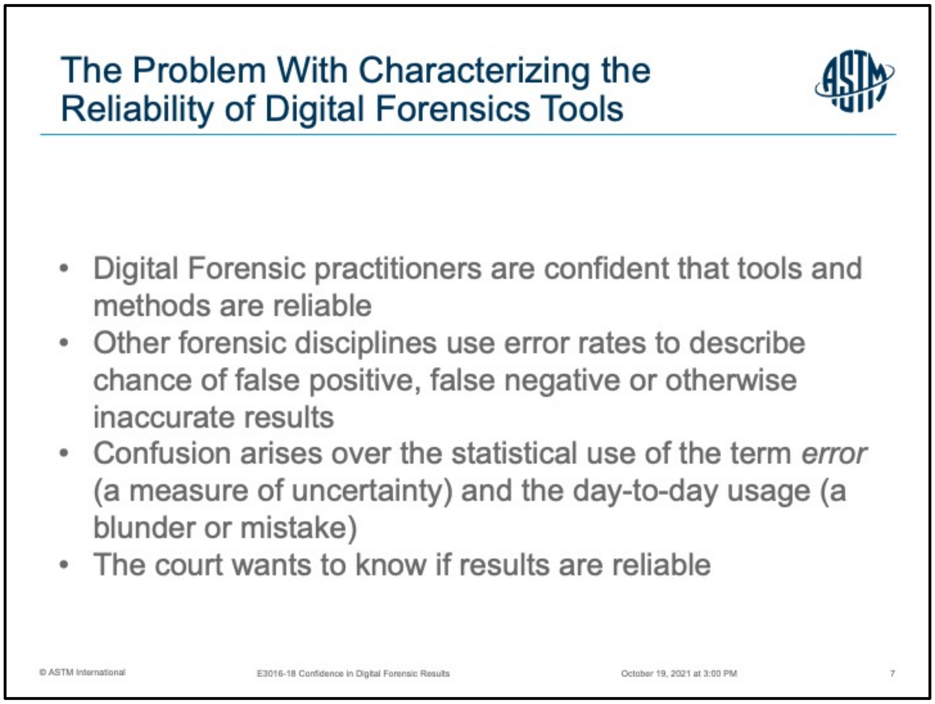
I’ll talk about some examples.




First we need some background for digital investigations.



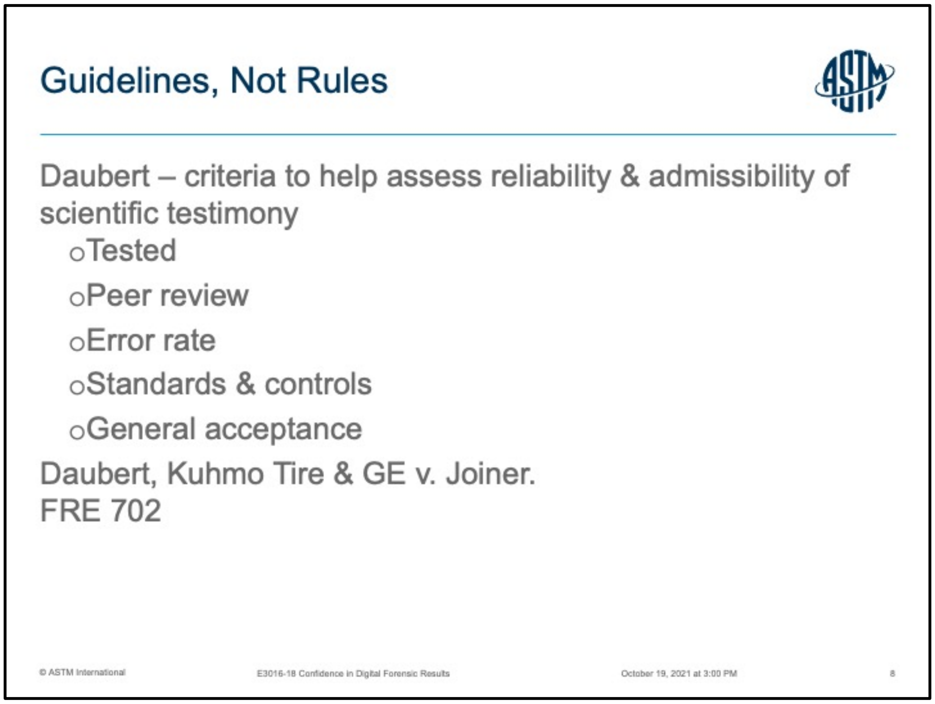

Lord Kelvin had a lot to say about what was science and how it ought to be done. What we need to do is measure reliability. Often some one will ask how reliable is what you do. In many cases if you can answer that you have an error rate then everything is fine. It shows that you understand the limits of your technique. But, as Lord Kelvin cautions in the second quote, don't over rely on the same measuring stick for everything.



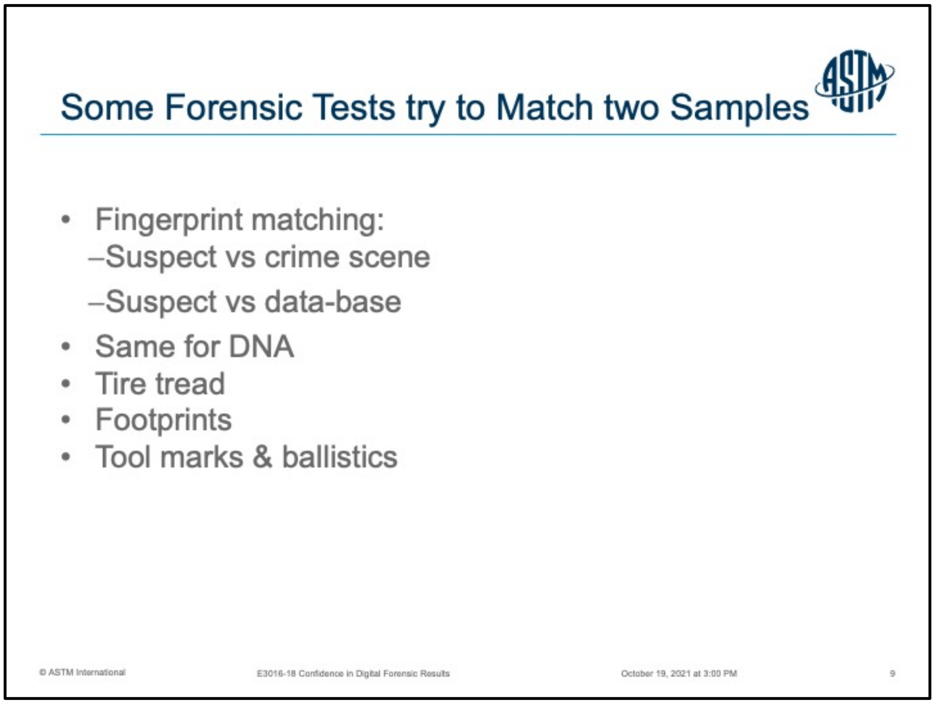

The court wants to know if results presented are reliable.

We know that our results are reliable, but how can we communicate this to the court.

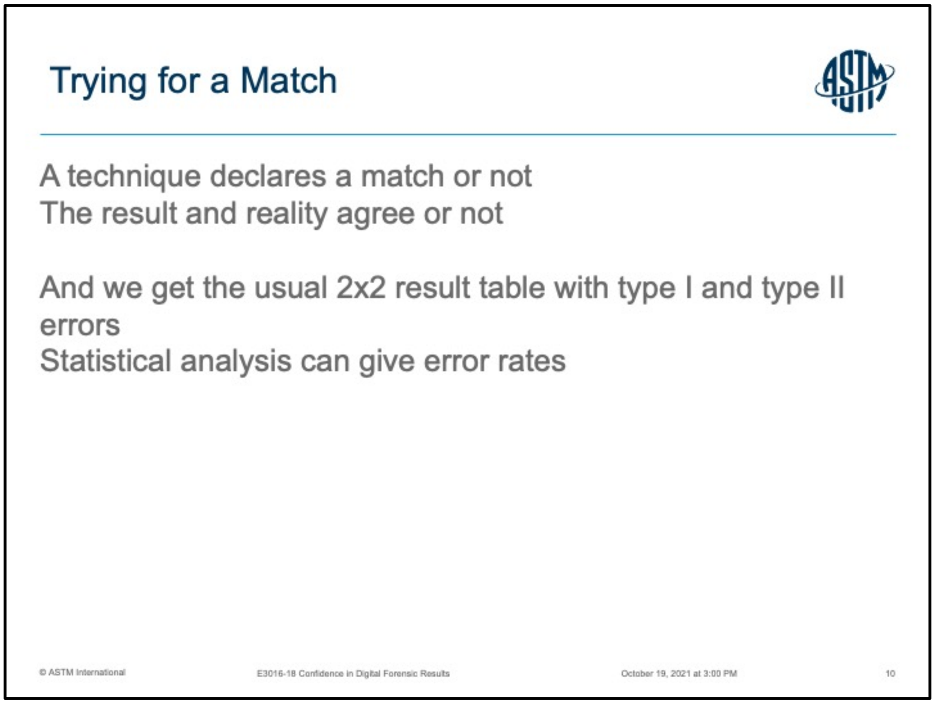
Other disciplines can often use error rates to describe the chance of false positives or false negatives or otherwise inaccurate results, but we do not always have that. The term error often causes a problem because the statistical meaning is a measure of uncertainty while the day-to-day usage is a blunder or mistake.




There are guidelines for reporting reliability of a technique, but Remember these are guidelines and not rules. It's nice to be able to meet all of them but you don't have to. However, this is not legal advice, always check with your attorney.



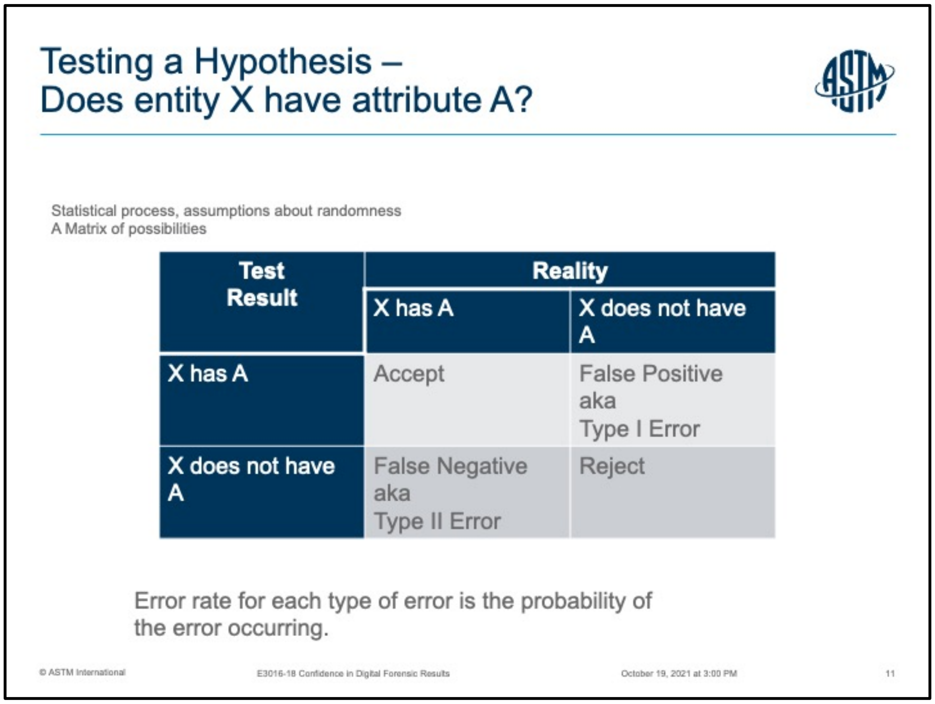

Other disciplines often focus on a single task such as matching one sample from the crime scene and a sample from a suspect. A simple straight forward question with a "yes" or "no" answer. Digital sometimes does this too, say to check if a suspect machine has any files from a set of known files that are of interest. Digital is not a single test, but many (dozens to hundreds) independent tests, that together form a narrative of events.




Trying for a match between two items has four possible outcomes, two that reflect reality and two that don't.

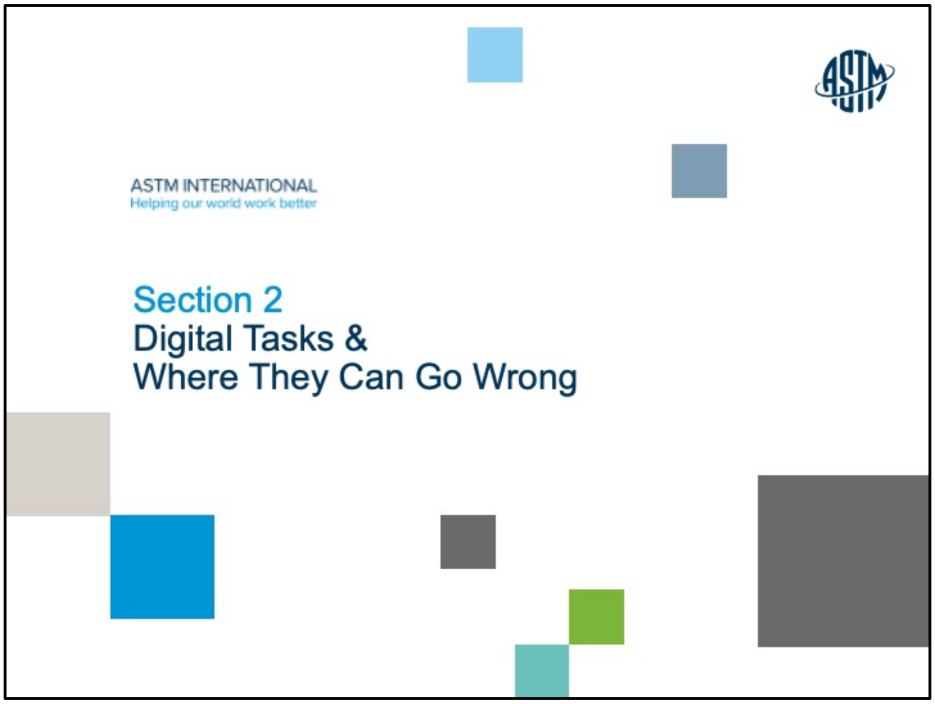
A test for matching two items is a natural task for using statistics to get error rates. The test reports either a match or not and the result is either correct or not.

Keep in mind that there is often an assumption (and requirement for valid statistics) that the population of test values follows a Normal, or in other words a Gaussian distribution.

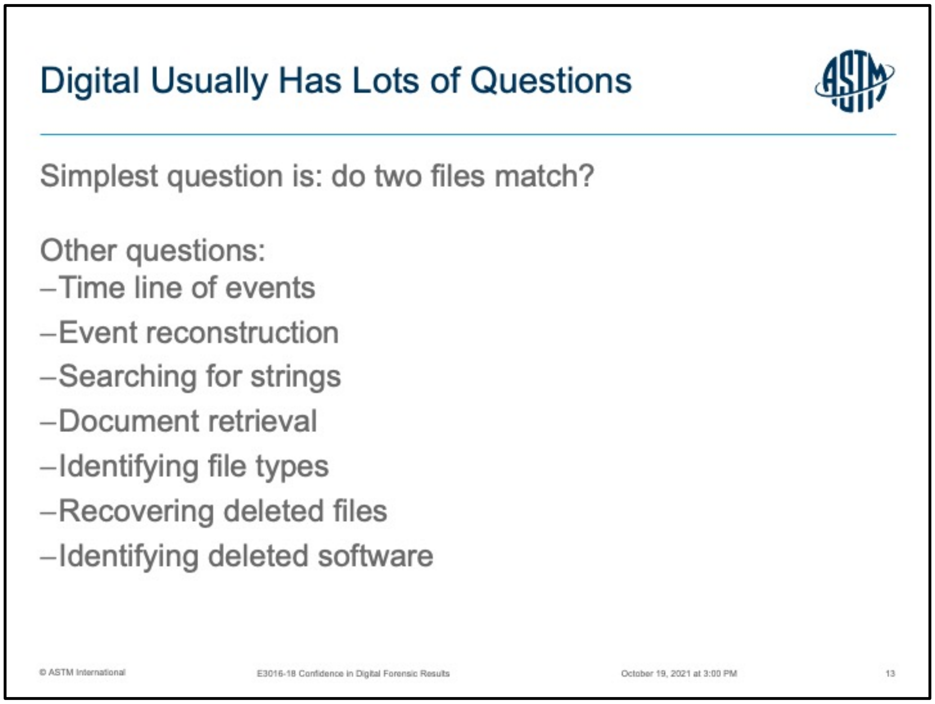



Matching is like a hypothesis test. Reliability can be measured with probability and then you can make statements about uncertainty. Some property is measured in each sample and then compared.

It is often tempting to use the average of the distribution, but this can give misleading results. For example, an error rate for a deleted file recovery tool might depend on some parameter like degree of file fragmentation and we could measure fragmentation of a large population of storage devices from SD cards to 5TB drives. The distribution of fragmentation rates across all the storage devices might show small devices have high rates of fragmentation and large drives have a small rate. The distribution likely looks like the two humps of a Bactrian camel, with the average falling in the valley between the two humps and would be misleading if used.




First we need some background about digital investigations.




As the investigator tries to assemble a narrative of events, there are many other unrelated tasks involved, each with varying risks to the reliability of the results.

## Digital-World vs Real-World

Digital is not as daunting as it seems!

| Correspondence of Real (non-digital) World to Digital World Evidence | |
|---|---|
| **Real-World** | **Digital-World** |
| **Crime scene or a place to search for evidence: could be a small site like an apartment or a large site like a farm or business.** | Computer, mobile device, storage device: a device to be examined; a server farm with many computers |
| **An item of evidence that is fragmented: shredded document, buried body** | Deleted data: evidence that isn't apparent with the usual computer user tools and can't be examined without some reassembly |
| **On site records such as a filing cabinet or desk** | Files stored on the computer hard drive, removable media. |
| **Offsite records such as at a business branch office, a summer home, or a storage locker** | Files stored on a cloud server, or off-line on removable media |
| **Burglar tools or weapons** | Hacking tools |
| **Names, phone numbers and addresses from a list of contacts, e.g., address book on paper.** | Contact list from a mobile device |

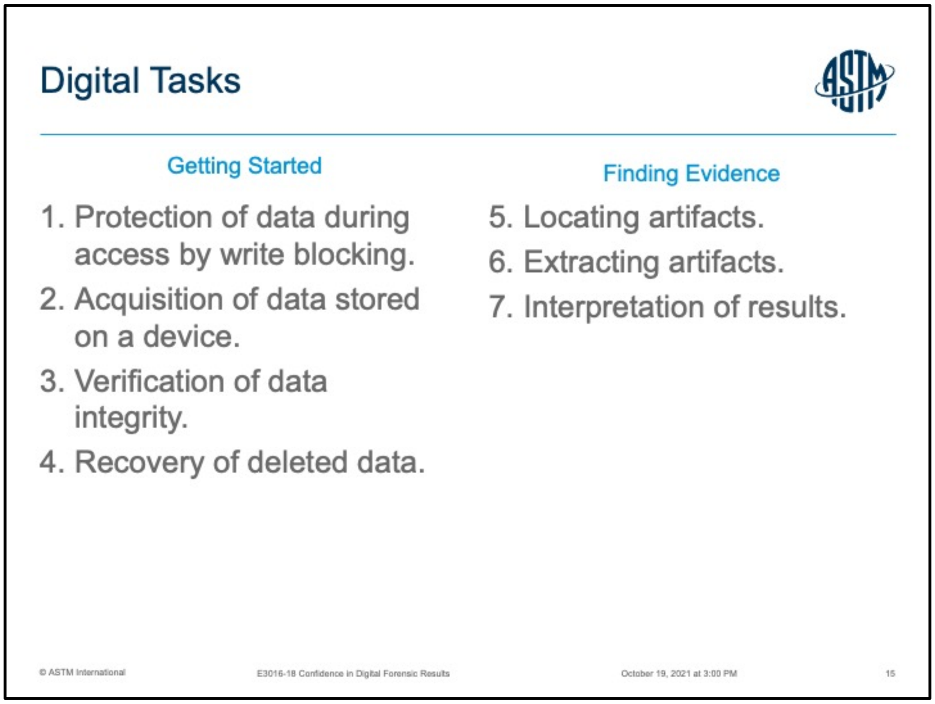



Most people are not computer experts even though almost everyone has to frequently interact with a computer. Digital evidence is often daunting at first, too many new terms, too much jargon, but a digital investigation isn't really very different from a not-digital investigation. Many concepts and digital objects have analogs in the real world.

Some differences are actually very convenient, such as a real world crime scene stays in place for a short wlile and then is cleaned up,

but you can make a copy of the digital crime scene (the digital data) and take it back to the lab.

Not just the items that caught your eye as you strolled through. At the lab you can revisit as often as you want.




You need to make an accurate copy of any relevant digital data without changing the original. If possible you want to acquire all the space on the storage device even if it is not currently used.

You can then examine the acquired data, but you may need to check that you don't accidently change anything.

The reason you want the unused space too is that computers are lazy and don't overwrite deleted data immediately and the deleted data can sometimes be recovered.

It is easy to search a digital file if you know what you want to find.

But then you have to understand what you got.

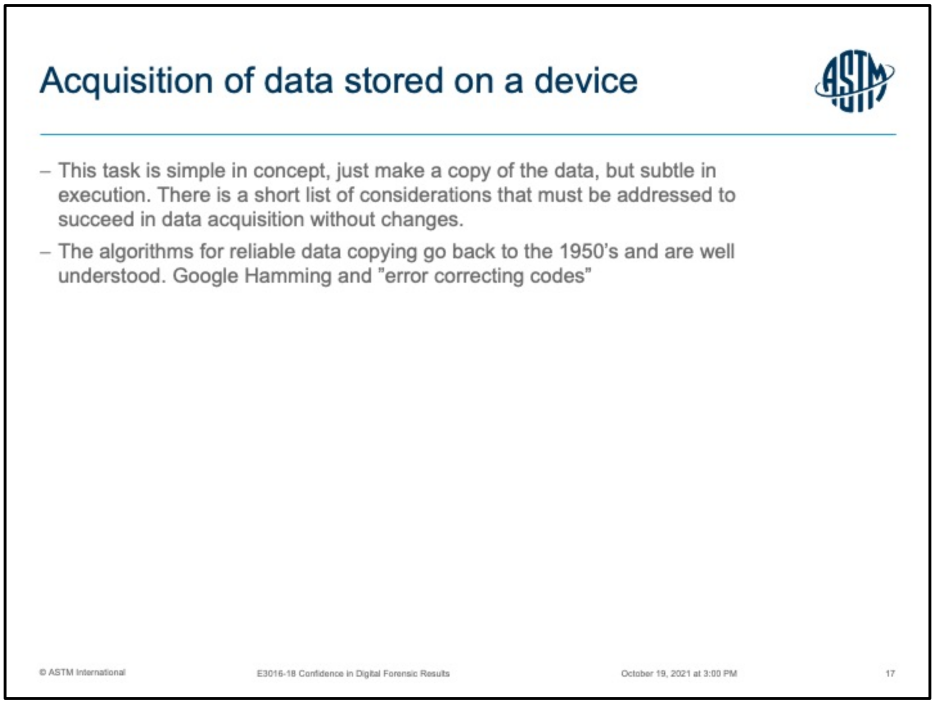
## Protection of data during access by write blocking

- Connecting a storage device to a computer may be necessary to acquire the data. If possible, techniques should be employed that do not allow any changes to the original data and allow the acquisition of the storage device contents accurately.
- Not always possible to use write blocking, sometimes a small program needs to be installed that overwrites some of the data to be acquired. This is often the case when acquiring computer memory. Sometimes the case when acquiring mobile device memory.



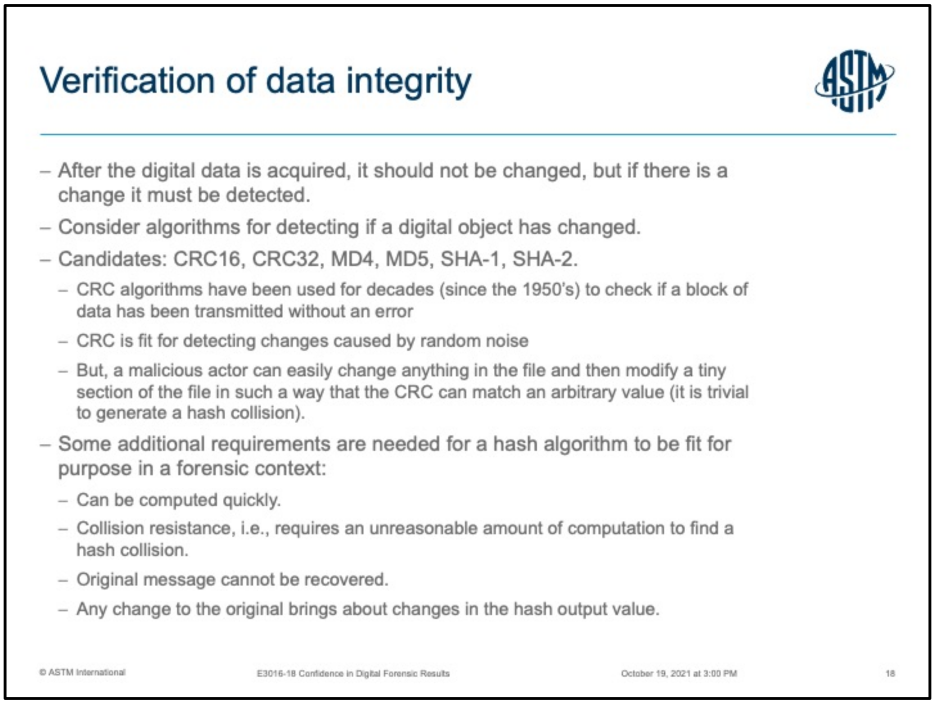
A hardware write blocker device is installed on the data/command path between a computer and a storage device. The blocker monitors all commands sent to the device and intercepts any commands that could change data contents on the device.

Software write blockers are also available.




This task is simple in concept, just make a copy of the data, but subtle in execution. There is a short list of considerations that must be addressed to succeed in data acquisition without changes.

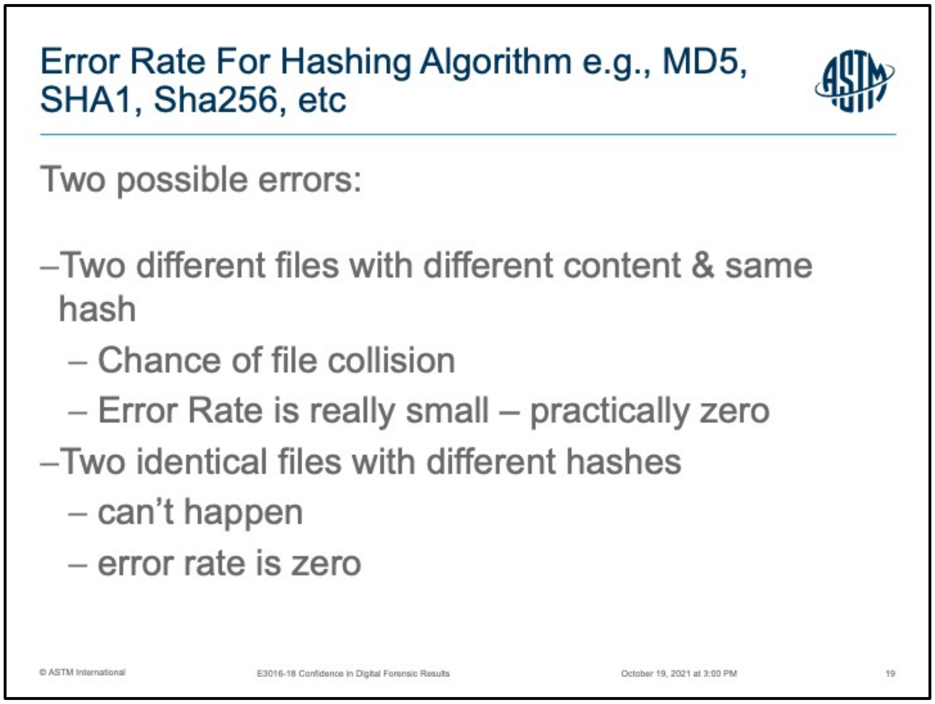
Copying data accurately is not a problem, but a tool may acquire the wrong data (you ask for user “john’s” files and you get “Natasha’s” instead), or  the device may have an unreadable area and has to return something. It just won't be something that was on the storage device.




The simple way to check if a working copy of a file has changed is to have a backup copy in addition to the working copy. The working copy can be examined and if there is any change it can be detected by comparison to the backup copy.

But, it may be inconvenient to devote all the storage space required to keep two copies of any acquired data. Instead, keep one copy and a checksum or hash that is just a small number (less than 200 digits or so) rather than GB or TB of extra data.

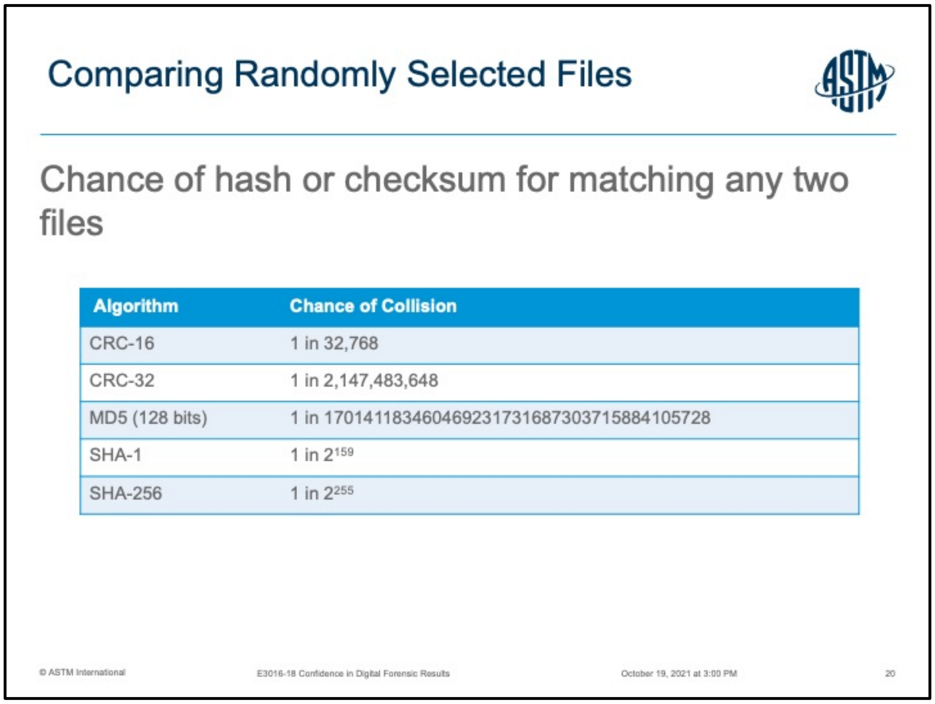
It is always possible for two unrelated files to have the same hash value, the more digits in the hash value to smaller the chance of a random "hash collision."




Hashing algorithms have a built-in chance of a false positive error that is unimaginably small. A false positive occurs if two files have the same hash value. It is always possible to occur, but so unlikely that it never occurs by chance.

A false negative occurs when two identical files have different hash values. If this seems to happen when two different programs compute hash values, then one of the programs is faulty.

The algorithm is immune to false negative errors, but an implementation can compute the wrong value.



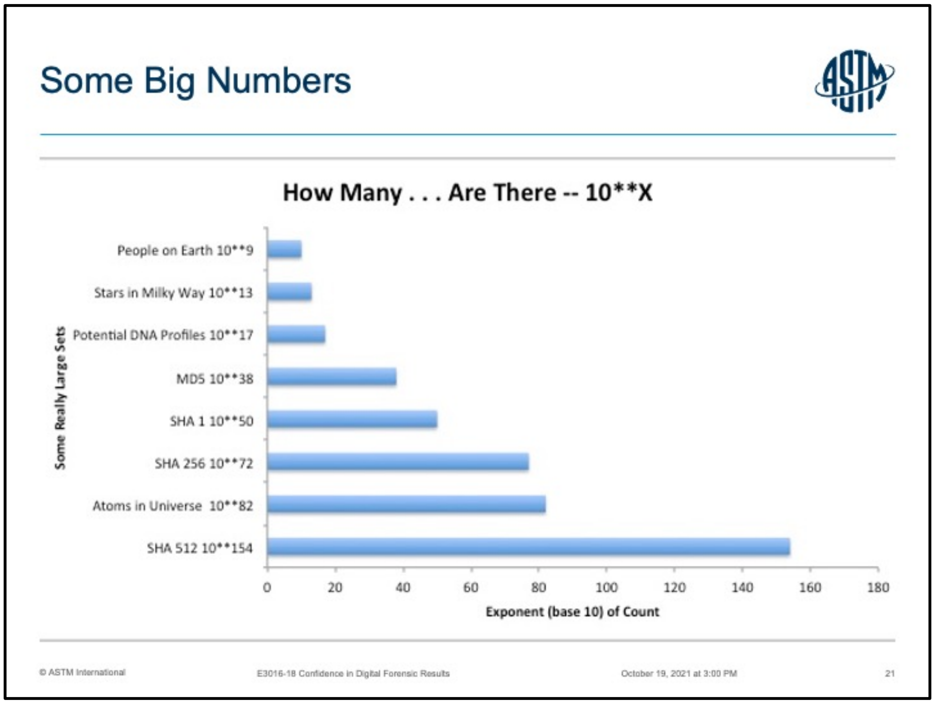

One in two billion looks like pretty good odds, why don't we use CRC-32?
You might be asked if you have validated your tool. Validation means "show that it is suitable for the task."

Now if you want to use CRC you might have looked at the CRC formula for CRC and tested your implementation to see if your tool calculates the expected values.

But even if all the calculations are correct you haven't validated the tool (shown it fit for purpose) you have only verified that the implementation is correct; you built the tool right.

Validation is showing that you built the right tool. You need to show that CFC meets additional criteria to be fit for purpose. Spoiler alert: CRC fails.

The CRC checksums lack some desirable properties of the cryptographic hash algorithms like randomization of the output so that CRC values for similar files might be similar, but for a cryptographic hash, similar files (even one bit different) produce very different cryptographic hashes.



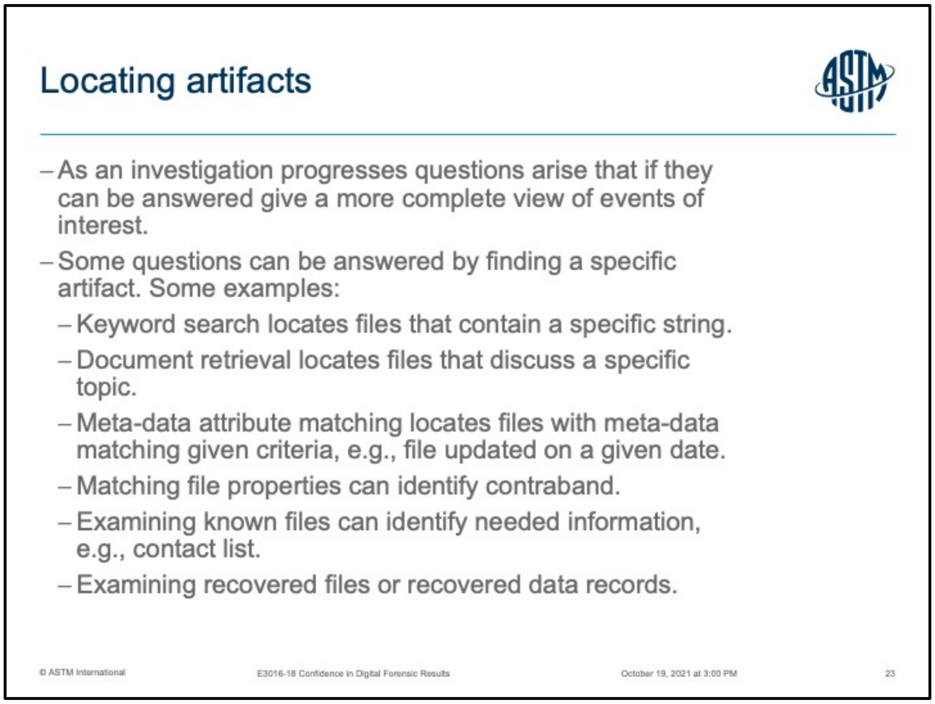

The probability of a hash collision is unimaginably small. MD5, considered “not good enough” by some, has a chance of hash collision better than one in the number of people that there would be if every star in the milky way galaxy had 10 planets with earth size populations (10**9 x 10**13 x 10 is only 10**23, this is far less than 10**38).

SHA512 is just overkill that’s been overkilled.

The objection to MD5 & SHA-1 is just making sure because Wang Xiaoyun (王小云) showed it is possible to create two different files with the same hash. This is a serious risk for some applications like a digital signature, but for most forensic applications such as verification that a file is unchanged, this is not a significant risk because of the limitations of her technique. While she can create two files that have the same hash value she can’t pick what the hash value is and the two files must be almost identical and can only differ by about 16 bits.

# Recovery of deleted data

- Data that has been deleted may be gone from access via the operating system, but the deleted data can be recovered with some caveats. Three types of data recovery are:
  - Meta-data based. Use remnants of information used to provide location data to partially reconstruct the deleted file. Some of the caveats are that the location data may be corrupt or file data may be overwritten.
  - File carving. There may not be any location remnants, but some files such as pictures or documents are highly structured and have identification codes at the beginning and the end of a file. After the file has been deleted, these codes can be found and the deleted file reconstructed. Similar caveats apply.
  - Deleted Record Recovery. Some files such as data bases are highly structured and frequently updated. Records (think of a line of data in a table) are created, updated or deleted. If the application leaves updated or deleted records in place they can be identified and retrieved.
- There is a lot of potential for misinformation; the investigators must their knowledge, skills and experience to examine the results of data recovery.




Keyword search tools usually offer as a basic function “search for files with the string you provide.” The tool then returns the names of files with the given string.

These tools often offer functions like “find files with social security numbers.”

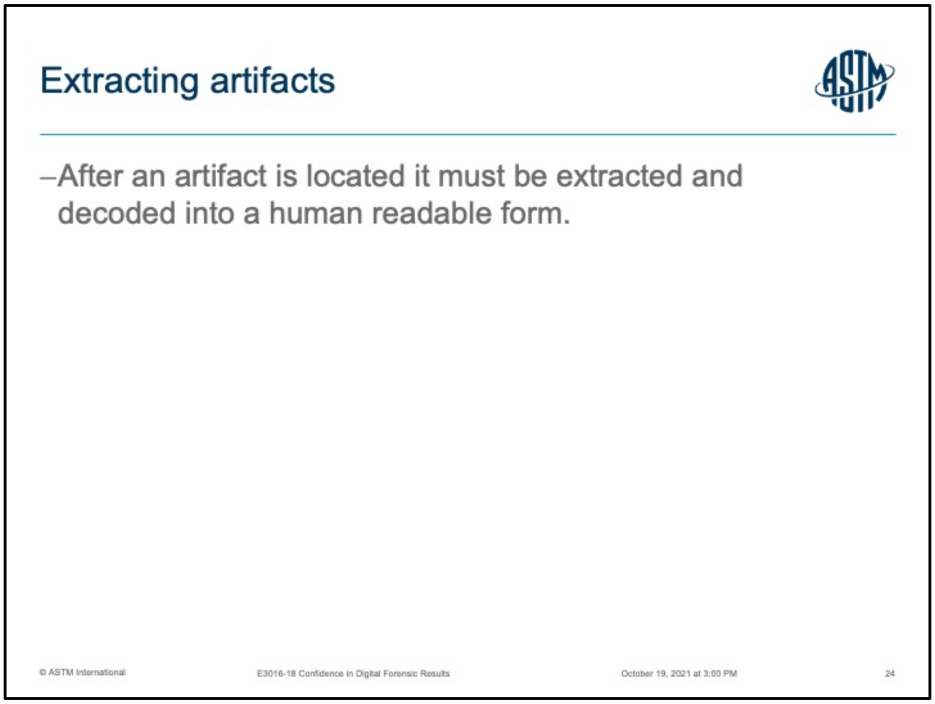
Another question to consider when testing a tool is to ask: does the algorithm the tool implements do what I want?

For example, string search tools often have a built-in feature to look for social security numbers. When we tested one string search tool, the tool offered two ways to do the search: live search and indexed search. We found that the two search methods gave different results. For a good reason, but the tool user should be aware.

It is possible to owe US income tax, but not owe social security tax. So you don't have a social security number. The IRS is very helpful and will give you an invalid social security number with a nine as the first digit to use as a tax payer ID number. Valid social security numbers never begin with an eight or nine.

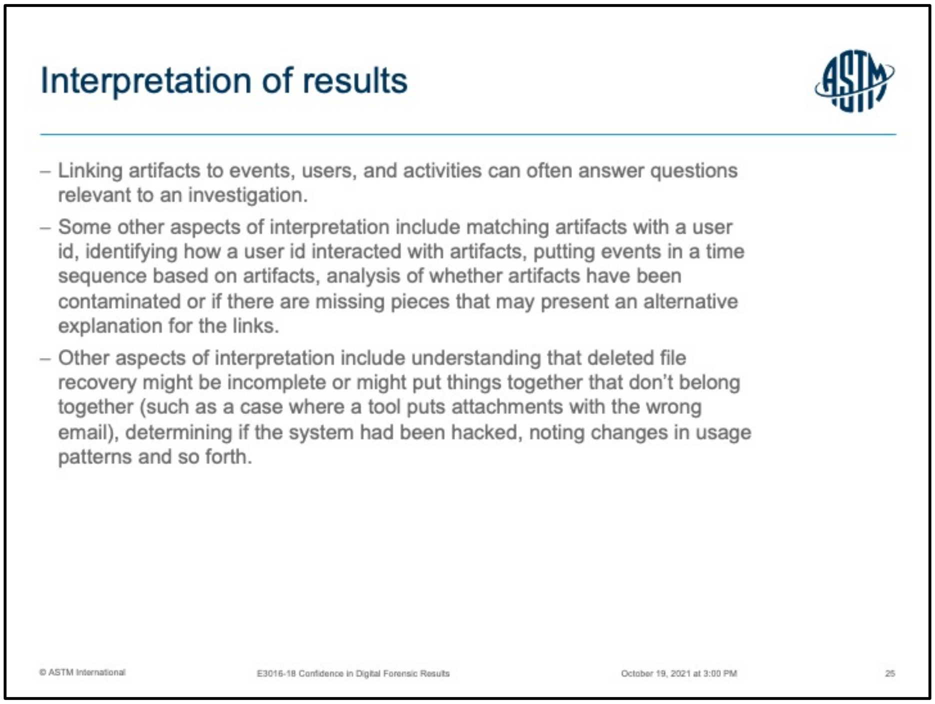
The indexed method looks for three digits, a hyphen, two digits, another hyphen and four digits, but the live search adds the criteria that if the first digit is an eight or nine, the string is not reported.

You need to know what the algorithm does, so you know if the tool addresses what you need.

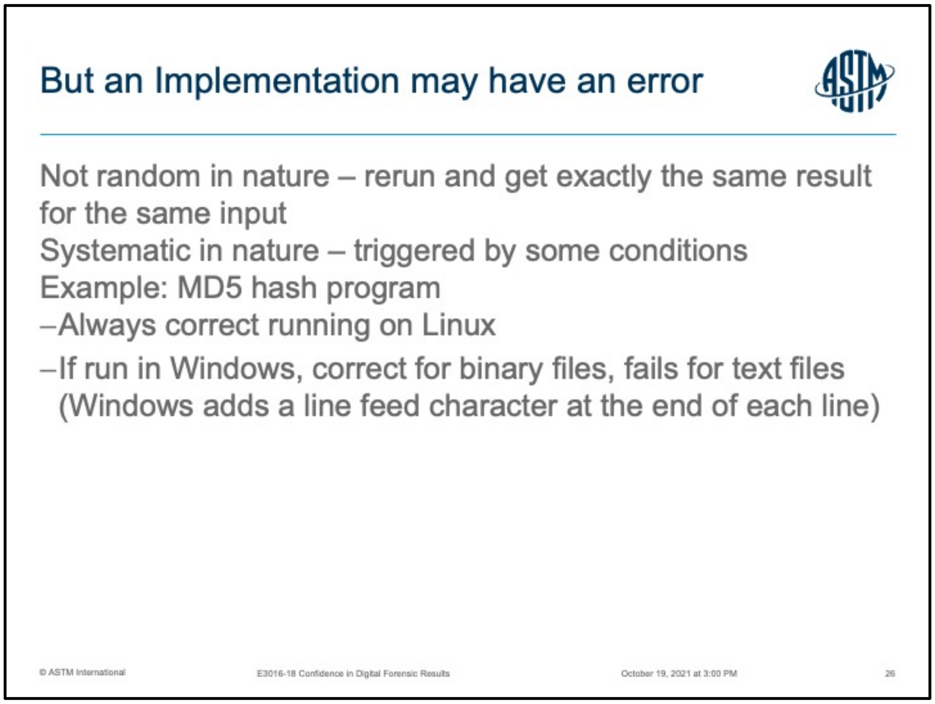



You have to know what the binary bits of the object are supposed to represent.

It could be a count of times a web site was visited or the time of the web site visit or the pixels of a picture or anything else that could be stored in  computer.



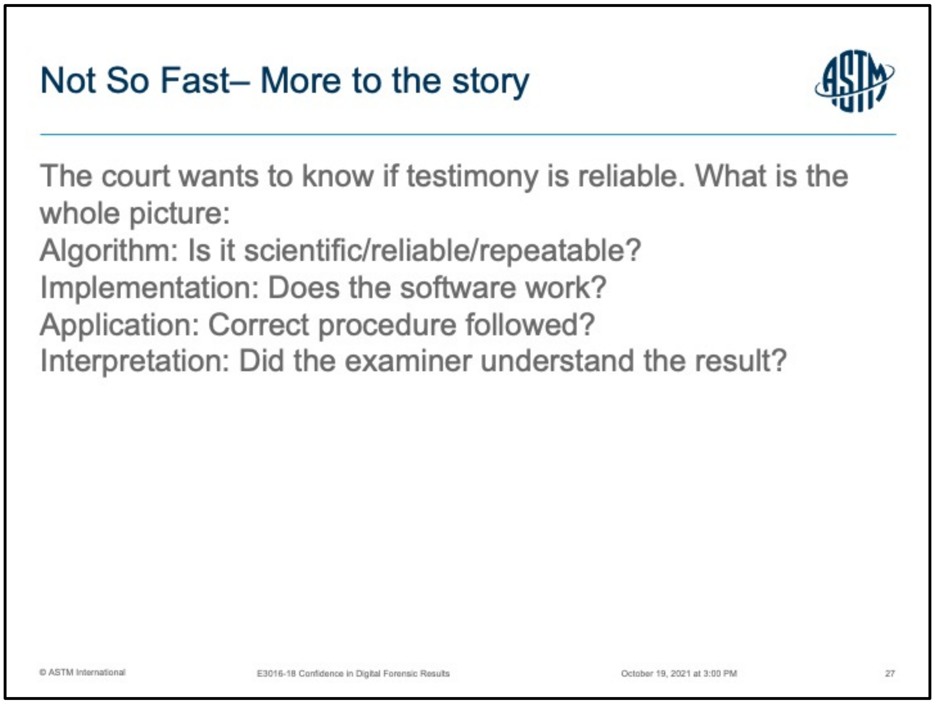

This is the critical step.




Here comes the rub, and it applies to any forensic process that uses computer software to calculate a result. A hypothesis test or a probability value depends on a random variable with a known probability distribution (usually Gaussian, aka Normal). The (random) error rate is a measure of uncertainty.

The software that makes the calculation can have a software error that is not random in nature. This is a systematic error, nothing random here. Same input yields same output. The intended formula of a calculation might be x+27, but if the program calculates x-27 the answer will be wrong every time.

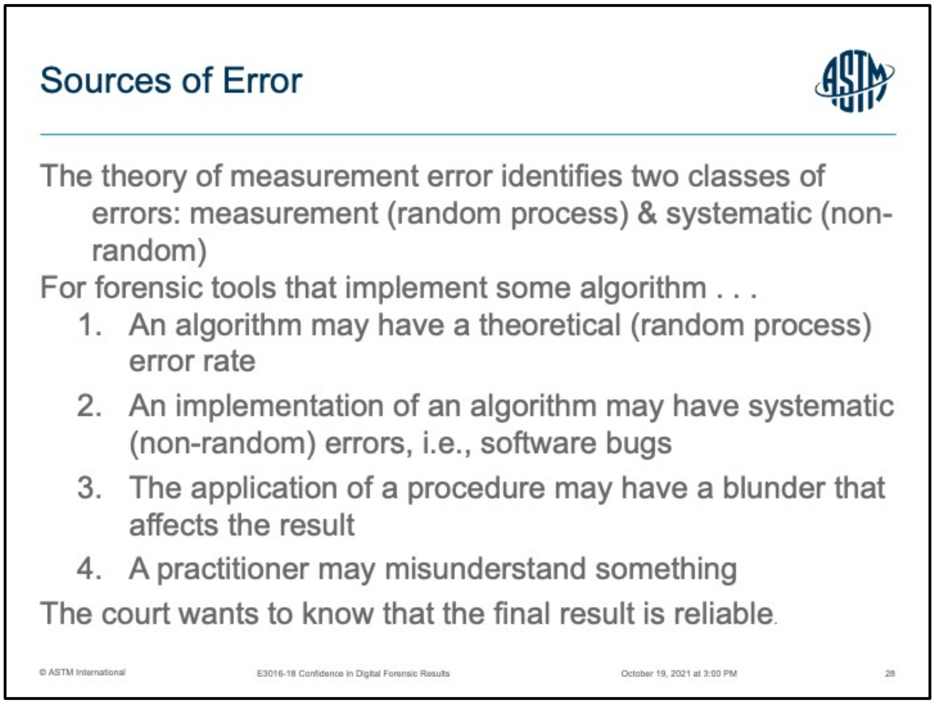
BTW, I wrote this program on Linux and moved the software to windows. The software error quickly showed up in just a few test cases and was promptly fixed.




An algorithm may have an error rate, but the tool implementing the algorithm may have systematic software errors and there are other broad paths to perdition.

A practitioner might not follow the best practice and wind up comingling data from two cases.

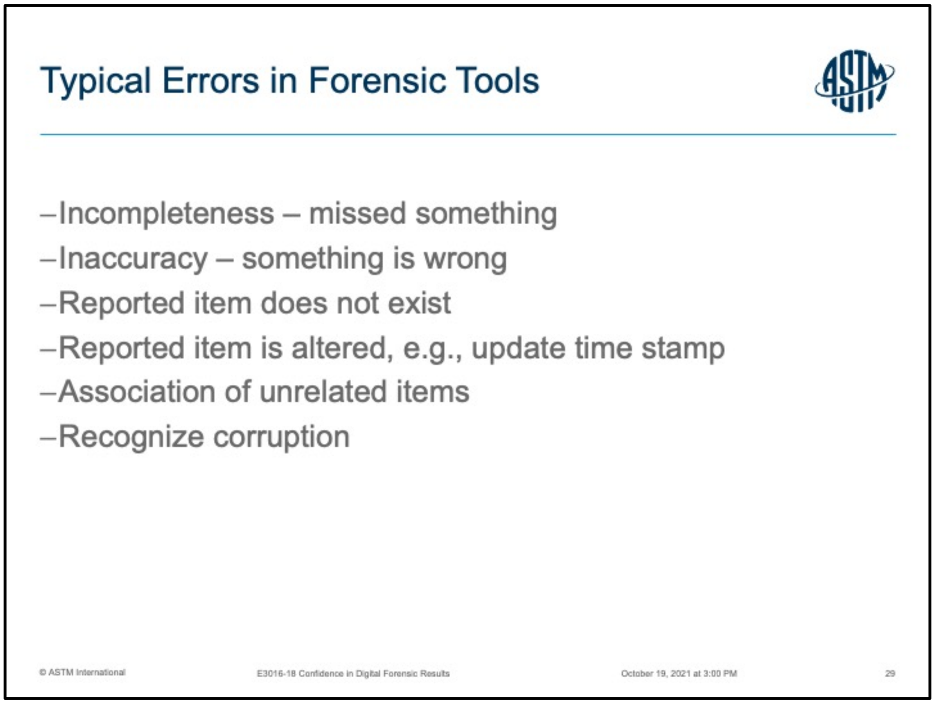
Or a practitioner may think that a file was accessed at 00:00 (midnight), but in reality it was zero because the “access” field was never updated by that particular OS.




Here is a little clarification on the word error

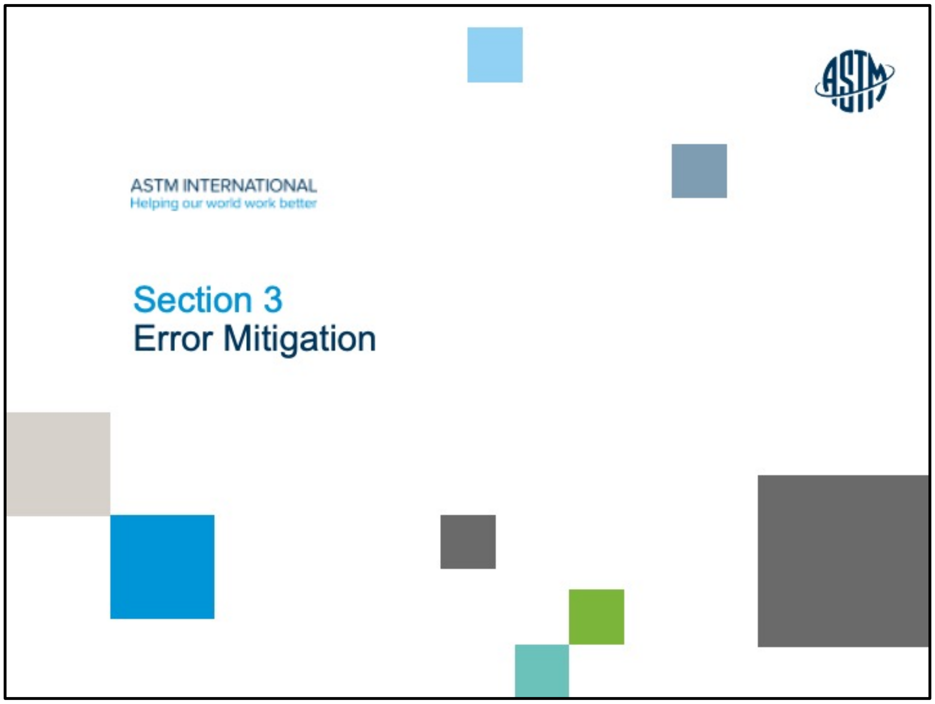
Statistical vs systematic

Again, the court wants to know the result is reliable




It helps to find errors if you can identify likely errors and then test for them.

These are the kinds of errors we have seen at the NIST Computer Forensic Tool Testing Project (CFTT) while testing digital forensic tools




First we need some background for digital investigations.

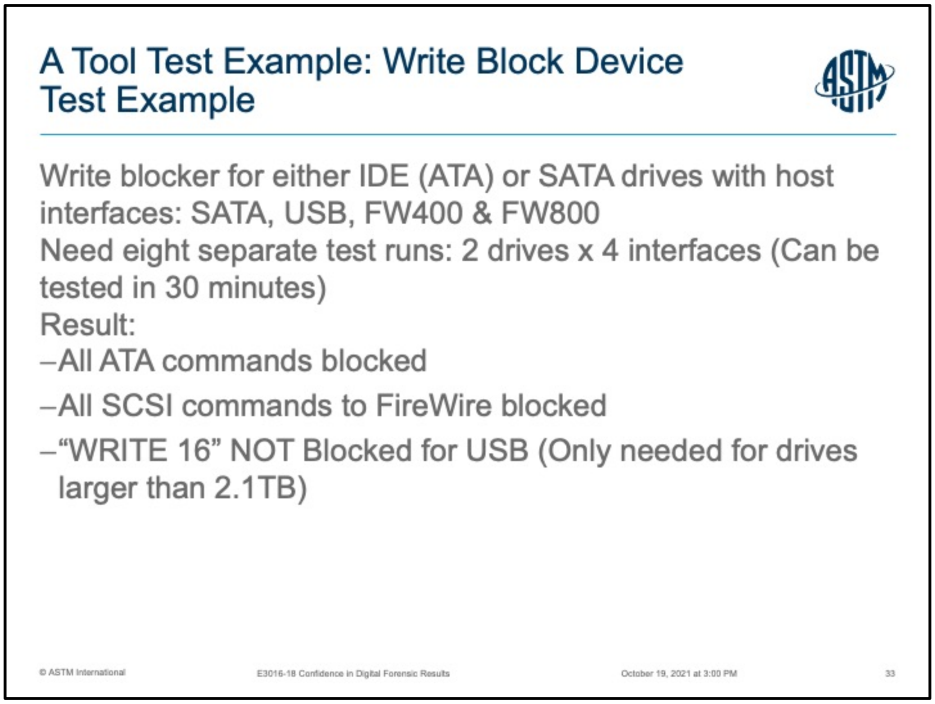
## Error Mitigation Strategies

- –Define likely errors & risks
- –Test tools for likely errors
- –Use written procedures
- –Document observations, history of problems
- –Oversight, Technical & Peer review
- –Context Analysis of results – sensible answer

## Three Examples of Error Mitigation Report

See These Examples in The Standard

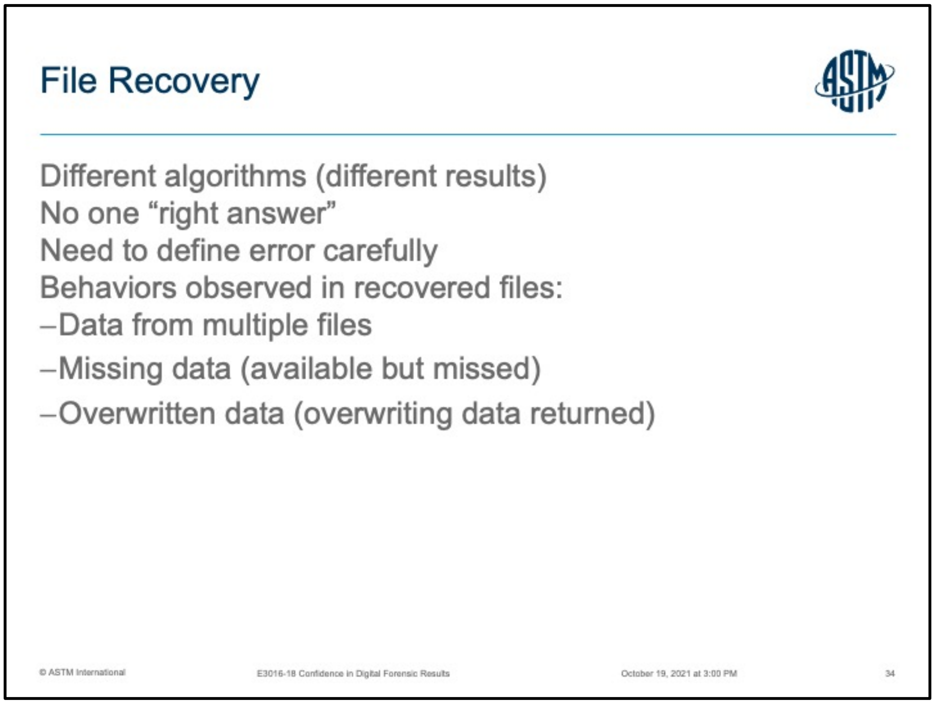
1. Intellectual Property Theft
2. New Technique Developed
3. Use of Tools Tested Elsewhere




Here is an example of what testing can reveal

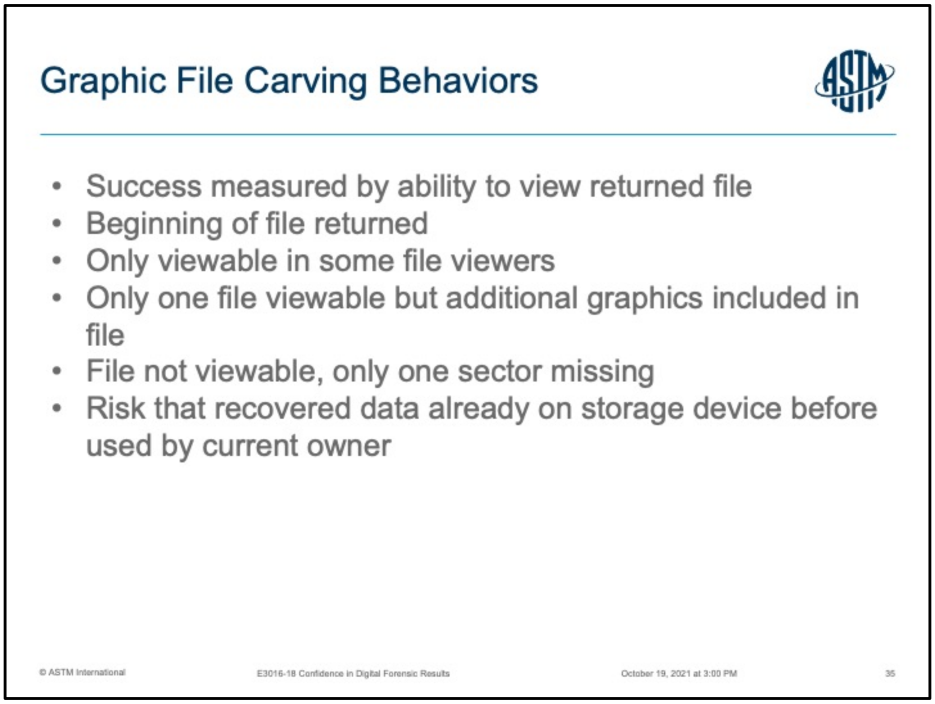
There are are about 5 write commands that a disk driver can choose from. A disk driver (software to access a storage device) usually has a preferred instruction for a given type of drive. In this case, on Windows XP, the write 10 command is preferred unless a disk address greater than 1.2TB is accessed. The “write 10” command has an address limit at that point and a different command, like “write 16”, with a larger address range must be used.

Note that this particular write block device works just fine except for “write 16” over the USB interface. “Write 16” is blocked on the firewire interface, but not the USB interface. The problem arose when a chip maker implemented, without informing the write block vendor, what from the chip maker’s perspective was a trivial change, but from the vendor’s perspective it was a significant change.




File recovery is one of the more challenging tasks. You need to test your tool so that you understand what results you can expect. Perfect file recovery is unlikely so you need to know what imperfections you might encounter.

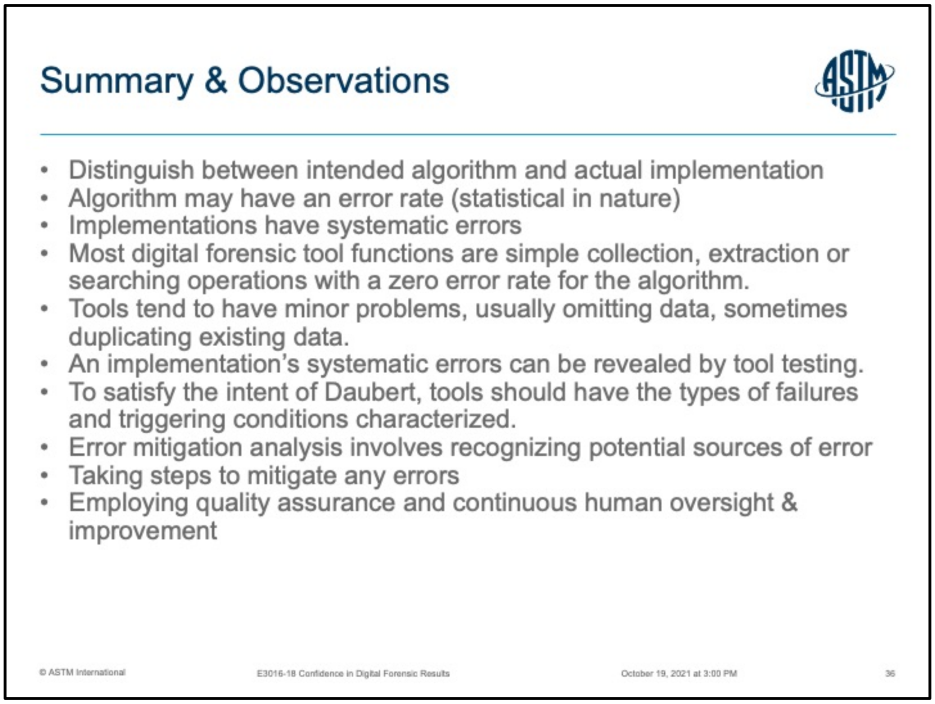
For example, recovered files need to be checked for mixing data clusters from multiple files together



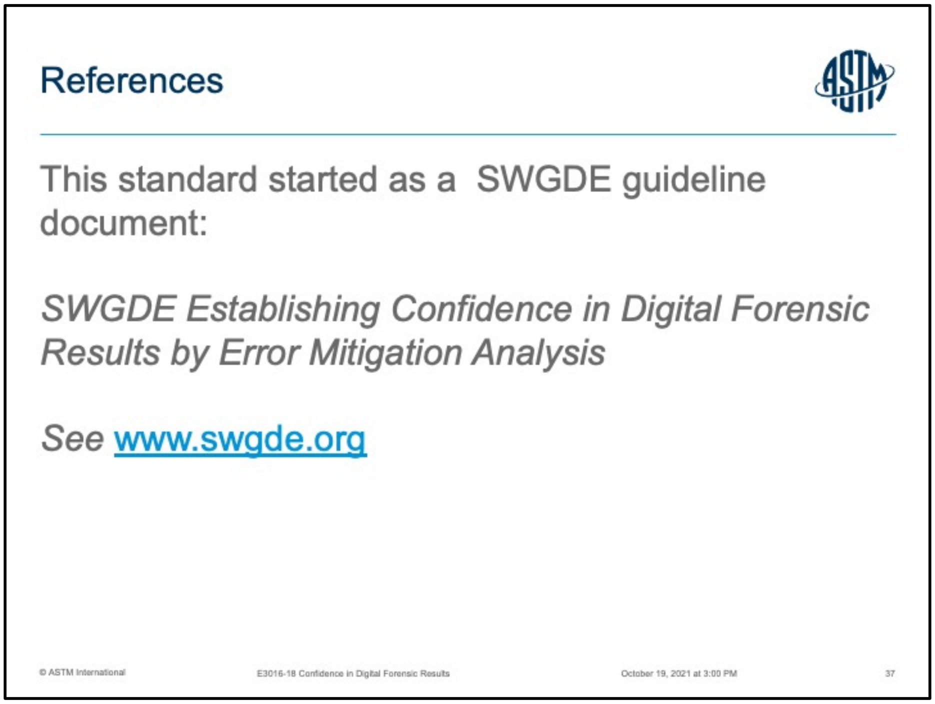

You can see a number of different behaviors with different file carving tools.

If you get a viewable result, the imperfections are often easy to identify.

Viewing a file usually makes any mixing of data from multiple sources stand out and easy to identify.




The key message from the Standard is to look at Error holistically – examine what kinds of errors can occur, which ones are likely. Then systematically take steps to address and reduce error and to describe where potential errors (especially the likely ones) remain.

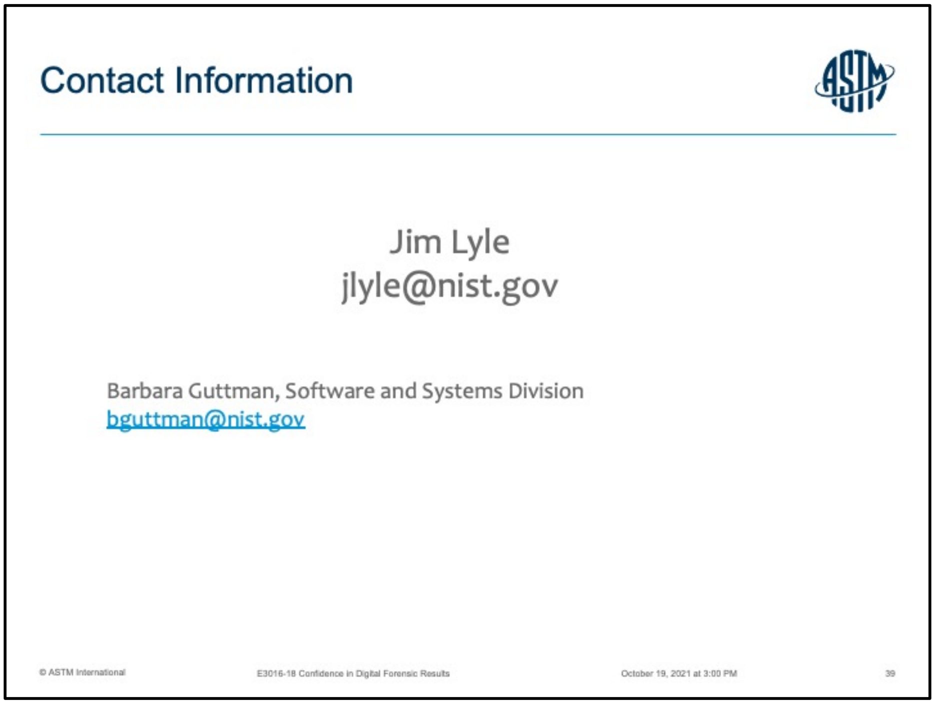



Here is a link to the SWGDE web site.

ASTM INTERNATIONAL
Helping our world work better

# Thank you

www.astm.org


Contact Information
Jim Lyle
jlyle@nist.gov
Barbara Guttman, Software and Systems Division
bguttman@nist.gov
© ASTM International
E3016-18 Confidence in Digital Forensic Results
October 19, 2021 at 3:00 PM
39