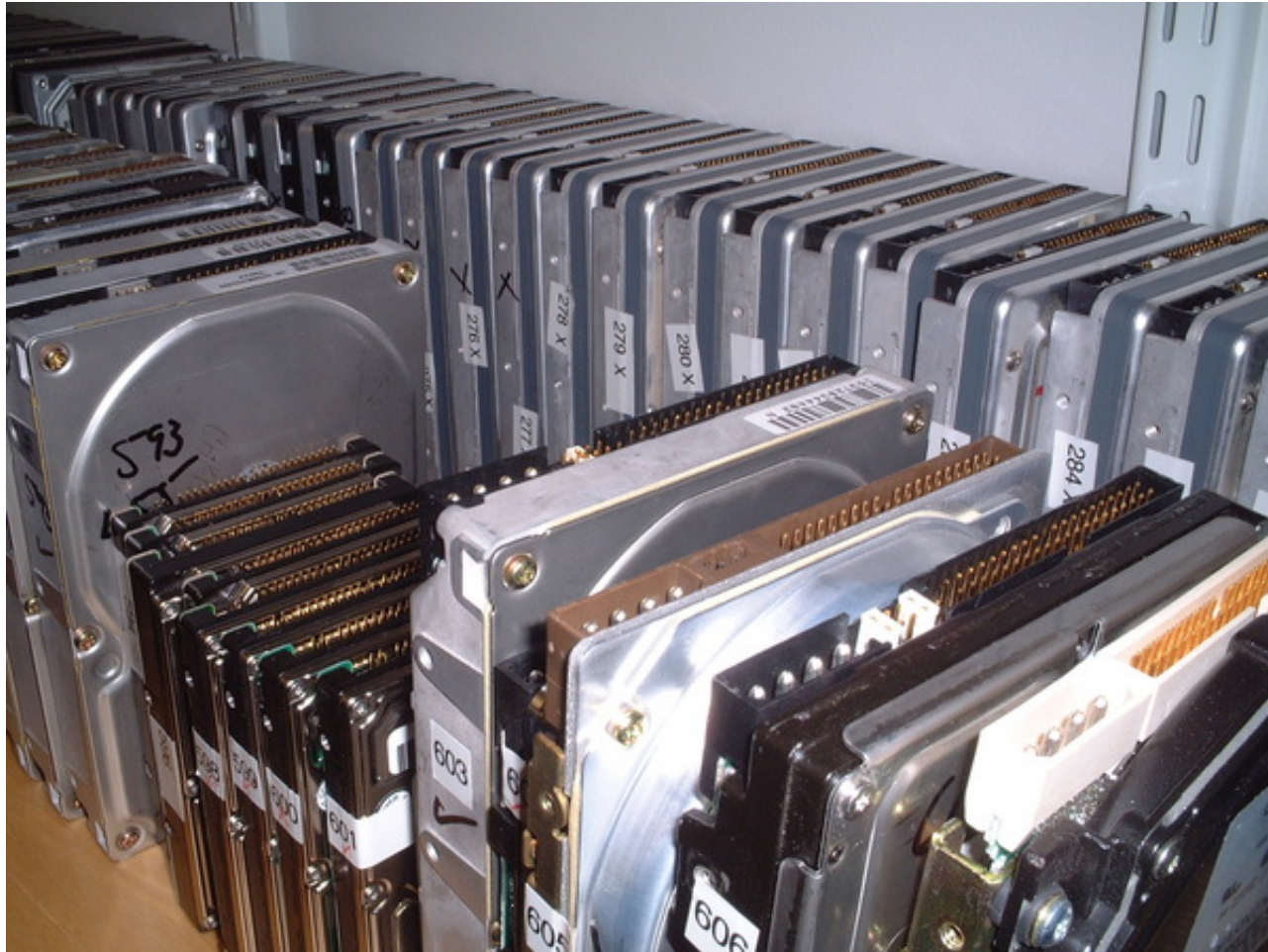


# Common-Mode Failures: What Can You Do With 236 750 Used Hard Drives?



**Simson L. Garfinkel**

Postdoctoral fellow, Center for Research on Computation and Society  
Harvard University

December 7, 2005

**Purchased used from a computer store in August 1998:**



## Computer #1: 486-class machine with 32MB of RAM

A law firm's file server...  
...with client documents!



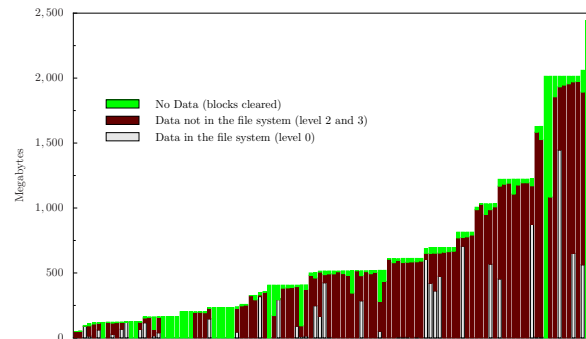
Computers #2 through #10 had:

- Mental health records
- Home finances
- Draft of a novel...

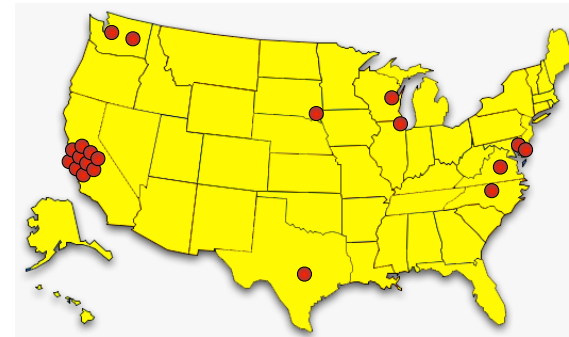
**Was this a chance accident or common occurrence?**

# This talk presents the disk sanitization problem and discusses a new technique for computer forensics.

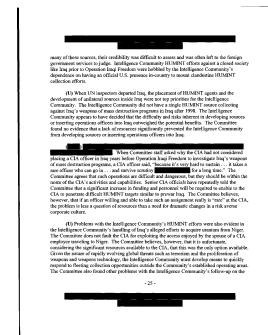
## 1. Scale of the problem



## 2. The Traceback Study



## 3. Common failures and solutions



# Hard drives pose special problem for computer security

Do not forget data when power is removed.

Contain data that is not immediately visible.

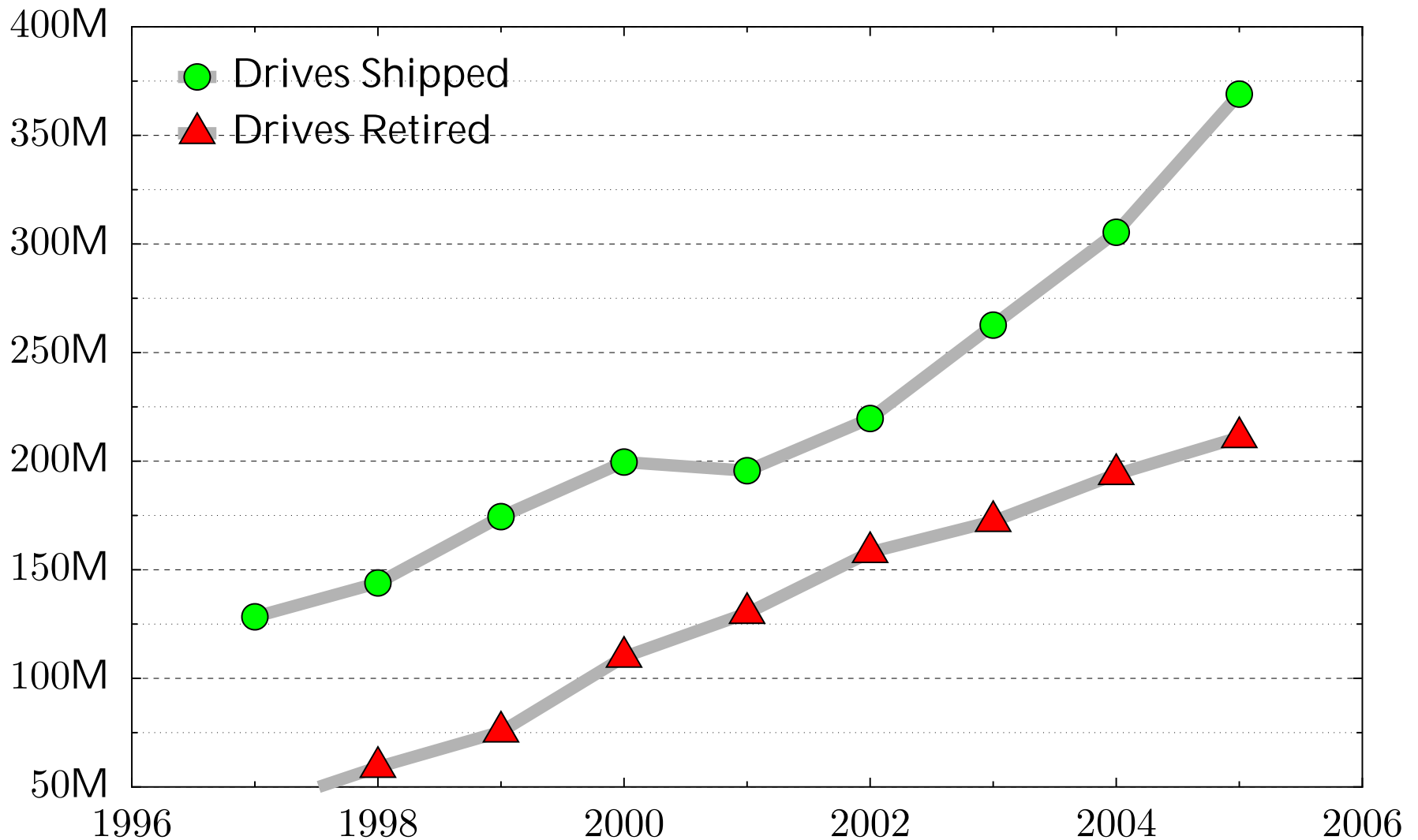
Today's computers can read hard drives that are 15 years old!

- Electrically compatible (IDE/ATA)
- Logically compatible (FAT16/32 file systems)
- Very different from tape systems



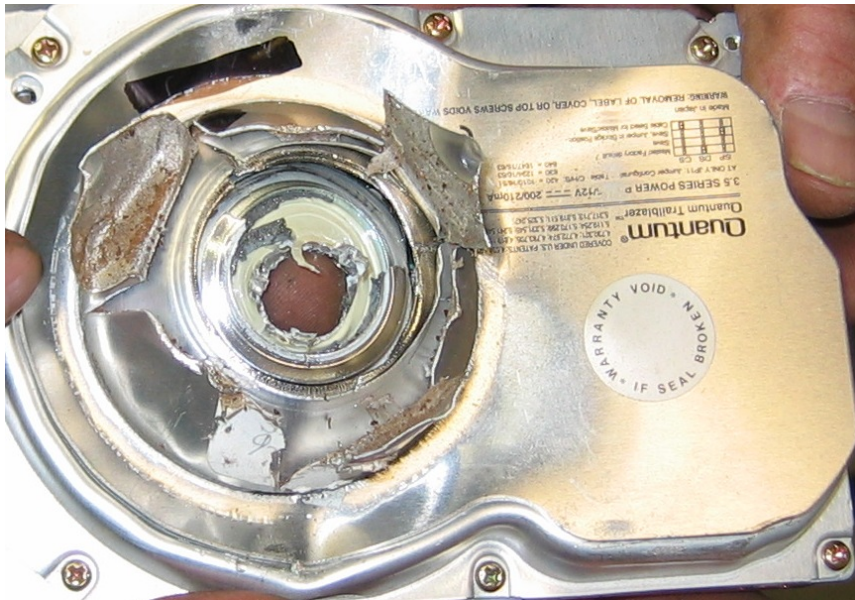


## Scale of the problem: huge!



**210 million drives will be retired this year.**

**Physical destruction will remove the information...**



**...but many “retired” drives are not physically destroyed.**

# There is a significant secondary market for used disk drives.



Retired drives are:

- Re-used within organizations
- Given to charities
- Sold at auction

**All Categories** [Save this search](#)  
350 items found for hard drives  
Sort by items: [ending first](#) | [newly listed](#) | [lowest priced](#) | [highest priced](#)

Picture Size	Item Title	Price	Bids	Time Left
	<a href="#">Lot of hard and floppy drives</a>	\$5.50	2	14m
	<a href="#">Lot of hard and floppy drives</a>	\$5.50	2	22m
	<a href="#">Lot of hard and floppy drives</a>	\$5.50	2	25m
	<a href="#">Lot of 2 hard drives IDE</a>	\$8.00	12	29m
	<a href="#">3.2 gig Hard Drives</a>	\$180.00	-	59m
	<a href="#">(5) 1.2 hard drives &amp; (15) 10/100 network</a>	\$15.00	1	1h 00m
	<a href="#">Lot of 3 Quantum 9.1 gig SCSI Hard Drives</a>	\$16.00	6	1h 25m
	<a href="#">IDE HARD DRIVES (3)</a>	\$6.50	6	1h 46m
	<a href="#">LOT OF 5 Hard Drives! 3.2 Gig Western Digital</a>	\$120.00 \$124.95 <del>78% off</del>	-	1h 50m
	<a href="#">QTY 3... IDE Hard Drives 2.5 Gg</a>	\$10.50	5	2h 02m
	<a href="#">5 WESTERN DIGITAL 2.5 GIG HARD DRIVES</a>	\$30.00	4	2h 03m
	<a href="#">QTY 3... IDE Hard Drives 1.0 Gg</a>	\$9.99	1	2h 04m
	<a href="#">Western Digital 850 meg IDE Hard Drives dutch</a>	\$6.00	1	2h 57m
	<a href="#">WINDOWS</a>	\$6.00	-	3h 18m

About 1000 used drives/day sold on eBay.



**Since January 1999,  
I have acquired 750 hard drives on the secondary market.**



## Drives arrive by UPS





# Data on drives “imaged” using FreeBSD





## Images stored on external firewire drives



**This is 900GB of storage.**

## For every drive, I catalog:

- Disk SN, date of manufacture, etc.
- Every readable sector on the drive..
- All visible files.
- MD5 of every file.
- MD5 of the image.





## Example: Disk #70: IBM-DALA-3540/81B70E32

Purchased for \$5 from a Mass retail store on eBay

Copied the data off: 541MB

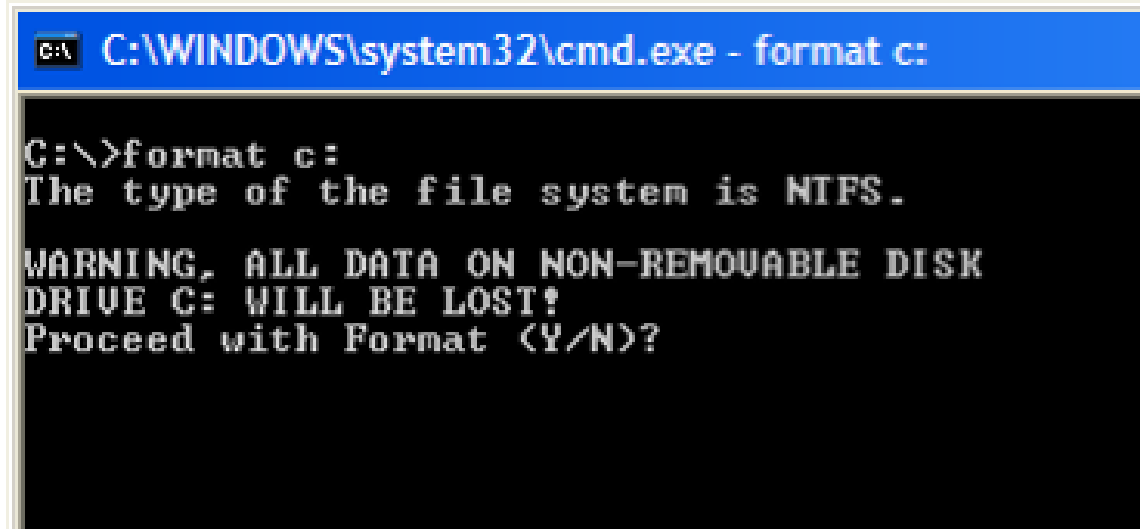
Initial analysis:

Total disk sectors:	1,057,392
Total non-zero sectors:	989,514
Total files:	3

The files:

drwxrwxrwx	0	root	0	Dec	31	1979	./
-r-xr-xr-x	0	root	222390	May	11	1998	IO.SYS
-r-xr-xr-x	0	root	9	May	11	1998	MSDOS.SYS
-rwxrwxrwx	0	root	93880	May	11	1998	COMMAND.COM

**Clearly, this disk had been FORMATED...**



```
C:\>format c:  
The type of the file system is NTFS.  
WARNING, ALL DATA ON NON-REMOVABLE DISK  
DRIVE C: WILL BE LOST!  
Proceed with Format (Y/N)?
```

**Windows FORMAT doesn't erase the disk...  
FORMAT just writes a new root directory.**

## UNIX “strings” reveals the disk’s previous contents...

Insert diskette for drive

and press any key when ready

Your program caused a divide overflow error.

If the problem persists, contact your program vendor.

Windows has disabled direct disk access to protect your lo

To override this protection, see the LOCK /? command for m

The system has been halted. Press Ctrl+Alt+Del to restart

You started your computer with a version of MS-DOS incompat

version of Windows. Insert a Startup diskette matching thi

OEMString = "NCR 14 inch Analog Color Display Enhanced SV

Graphics Mode: 640 x 480 at 72Hz vertical refresh.

XResolution = 640

YResolution = 480

VerticalRefresh = 72

## 70.img con't...

ling the Trial Edition

-----  
IBM AntiVirus Trial Edition is a full-function but time-limited evaluation version of the IBM AntiVirus Desktop Edition product. You may have received the Trial Edition on a promotional CD-ROM, a single-file installation program over a network. The Trial Edition is available in seven national languages, and each language is provided on a separate CC-ROM or as a separate installation program.

EAS.STCm

EET.STC

ELR.STCq

ELS.STC

## 70.img con't...

MAB-DEDUCTIBLE

MAB-MOOP

MAB-MOOP-DED

METHIMAZOLE

INSULIN (HUMAN)

COUMARIN ANTICOAGULANTS

CARBAMATE DERIVATIVES

AMANTADINE

MANNITOL

MAPROTILINE

CARBAMAZEPINE

CHLORPHENESIN CARBAMATE

ETHINAMATE

FORMALDEHYDE

MAFENIDE ACETATE



**[Garfinkel & Shelat 03] established the scale of the problem.**

We found:

- Thousands of credit card numbers
- Financial records
- Medical information
- Trade secrets
- Highly personal information



**We did not determine why the data had been left behind.**

## There are roughly a dozen documented cases of people purchasing old PCs and finding sensitive data.

- A woman in Pahrump, NV bought a used PC with pharmacy records [Markoff 97]
- Pennsylvania sold PCs with “thousands of files” on state employees [Villano 02]
- Paul McCartney’s bank records sold by his bank [Leyden 04]
- O&O Software GmbH – 200 drives.[O&O 05]



**None of these cases are scientifically rigorous.**

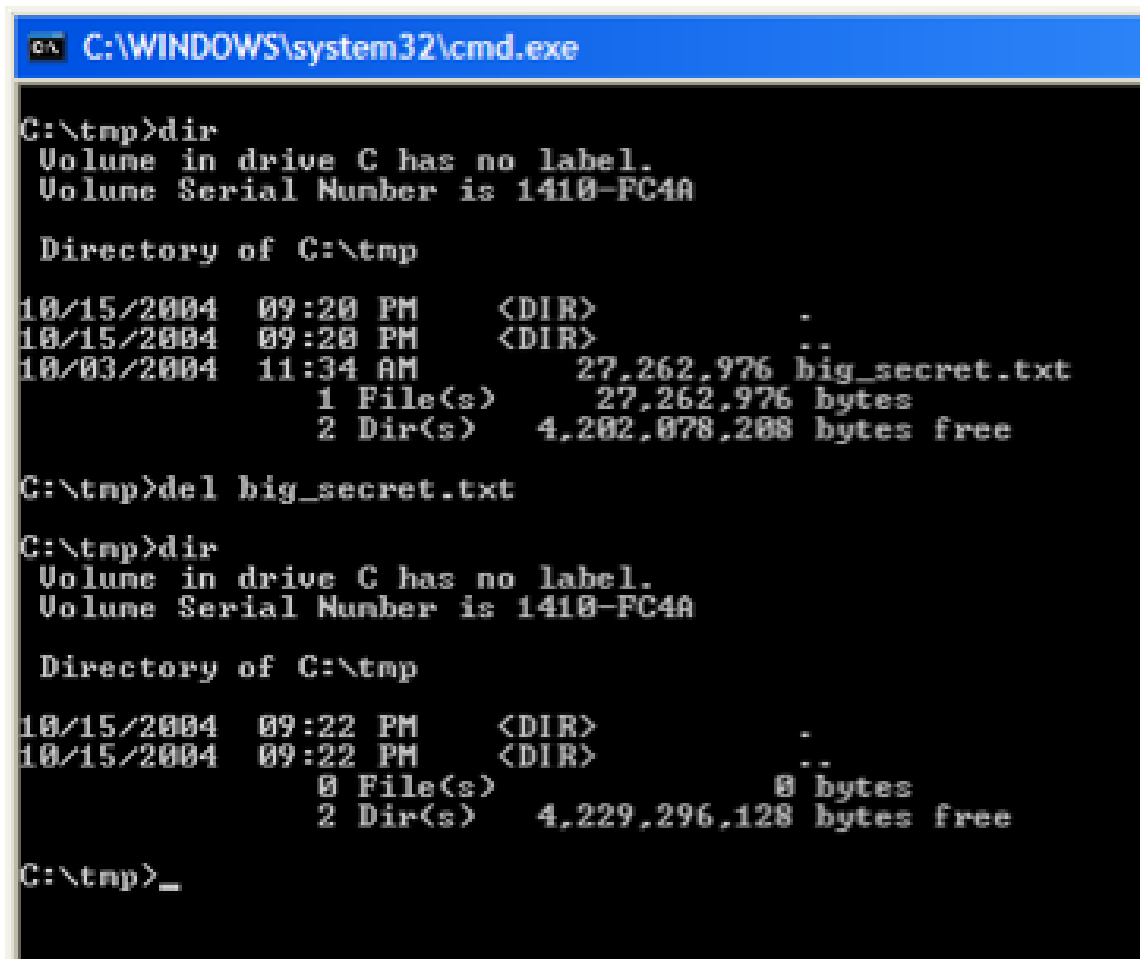
## Why don't we hear more stories?

Hypothesis #1: Disclosure of “data passed” is exceedingly rare because most systems are properly cleared.

Hypothesis #2: Disclosures are so common that they are not newsworthy.

Hypothesis #3: Systems aren't properly cleared, but few people notice the data.

## How could people not notice the data?



```
C:\WINDOWS\system32\cmd.exe

C:\tmp>dir
Volume in drive C has no label.
Volume Serial Number is 1410-FC4A

Directory of C:\tmp

10/15/2004  09:20 PM    <DIR>          .
10/15/2004  09:20 PM    <DIR>          ..
10/03/2004  11:34 AM             27,262,976 big_secret.txt
               1 File(s)              27,262,976 bytes
               2 Dir(s)    4,202,078,208 bytes free

C:\tmp>del big_secret.txt

C:\tmp>dir
Volume in drive C has no label.
Volume Serial Number is 1410-FC4A

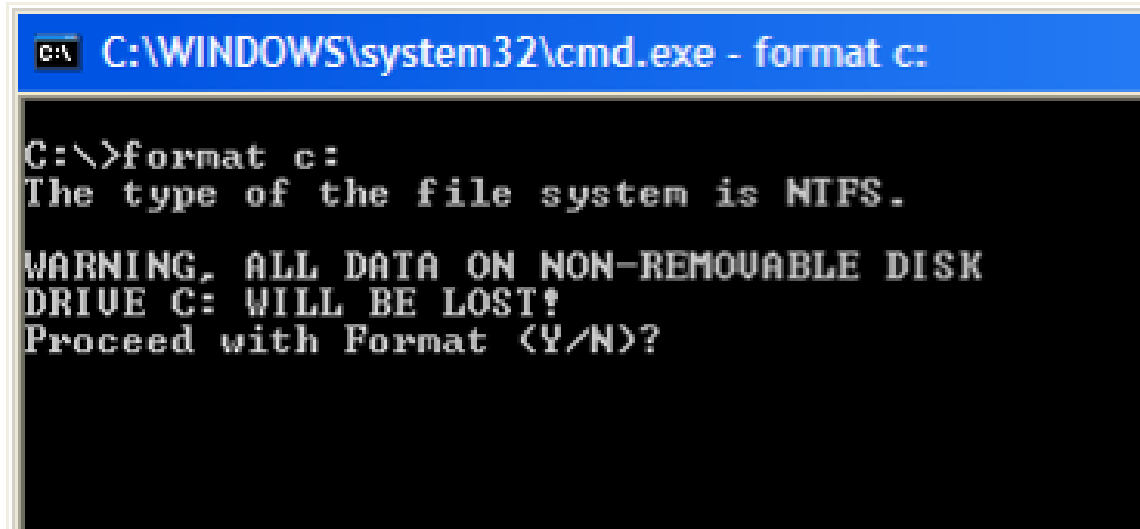
Directory of C:\tmp

10/15/2004  09:22 PM    <DIR>          .
10/15/2004  09:22 PM    <DIR>          ..
               0 File(s)                0 bytes
               2 Dir(s)    4,229,296,128 bytes free

C:\tmp>_
```

**DEL removes the file's name; doesn't delete the data.**

**FORMAT writes a new root directory and FAT.**



```
C:\WINDOWS\system32\cmd.exe - format c:

C:\>format c:
The type of the file system is NTFS.

WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
Proceed with Format (Y/N)?
```

**FORMAT doesn't doesn't overwrite the disk sectors.**



# Data left behind on hard drives is a serious social problem.

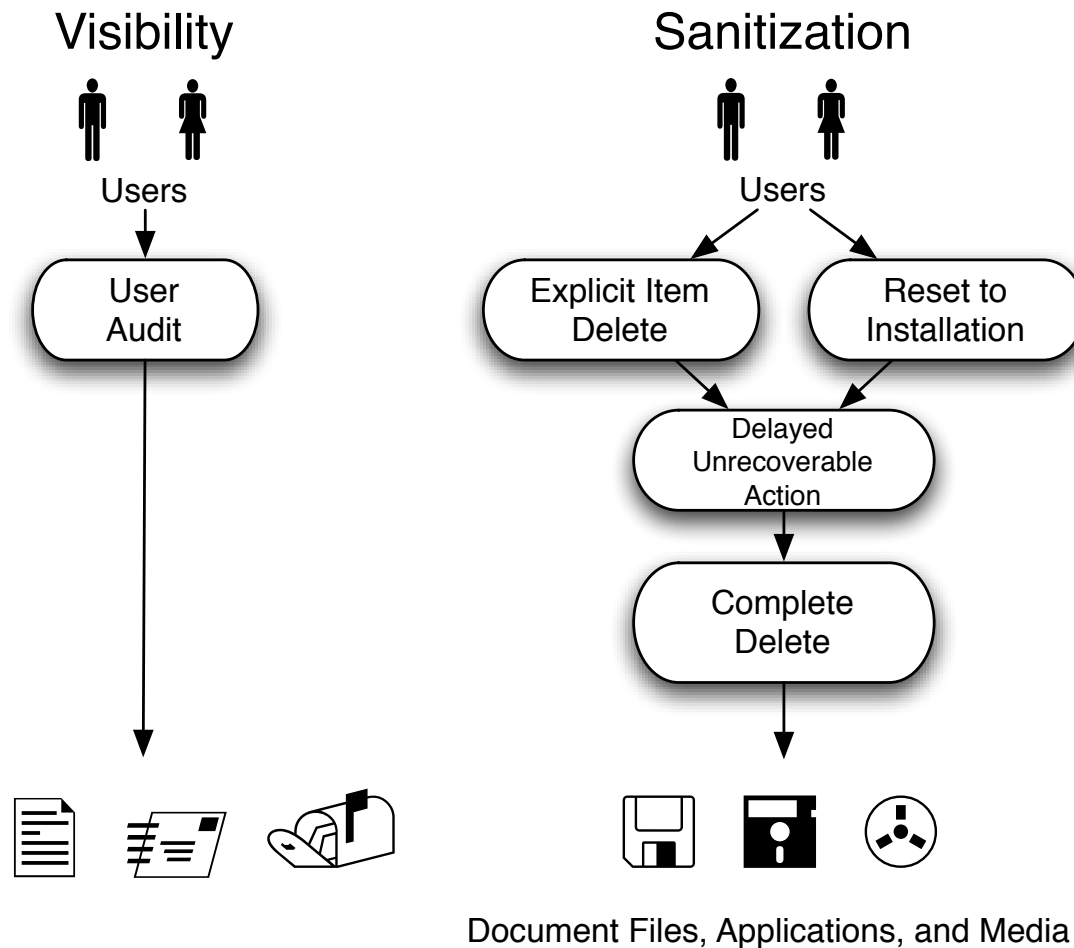
Large numbers of drives are being sold and given away.

Many appear to have hidden confidential information.



**We are morally obligated to solve this problem!**

## [Garfinkel '05] presents five distinct patterns for addressing the sanitization problem



<http://www.simson.net/thesis/>

## To be effective, a solution must address the root cause

### *Usability Problem:*

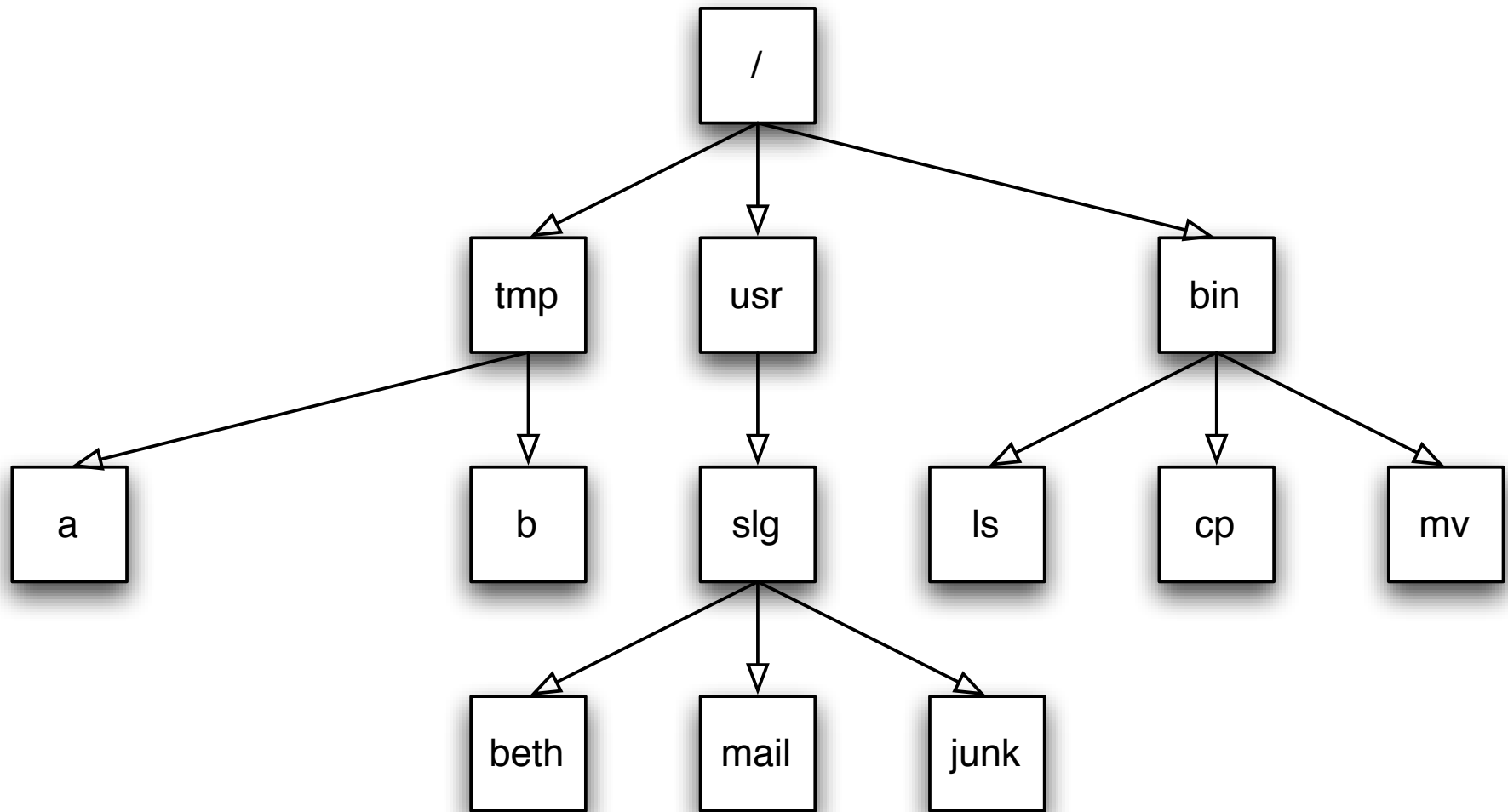
- Effective audit of information present on drives.
- Make DEL and FORMAT actually remove data.  
[Bauer & Priyantha 01]
- Provide alternative strategies for data recovery.

### *Education Problem:*

- Add training to the interface.  
[Whitten 04]
- Regulatory requirements.  
[FTC 05, SEC 05]
- Legal liability.

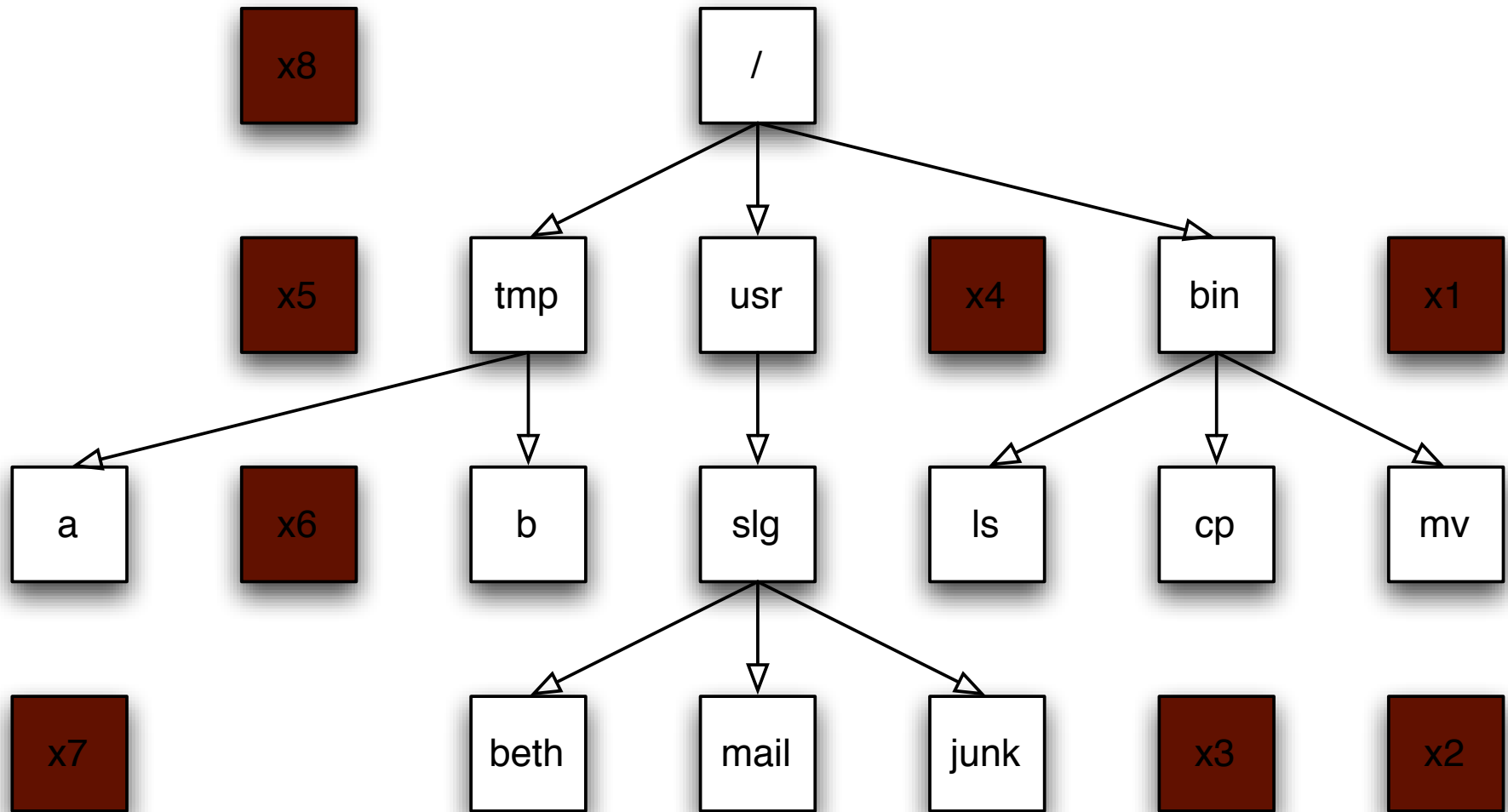
**To find that cause,  
I looked *on the drives* and *contacted the data subjects*.**

**Data on a hard drive is arranged in sectors.**



**The white sectors indicate directories and files that are visible to the user.**

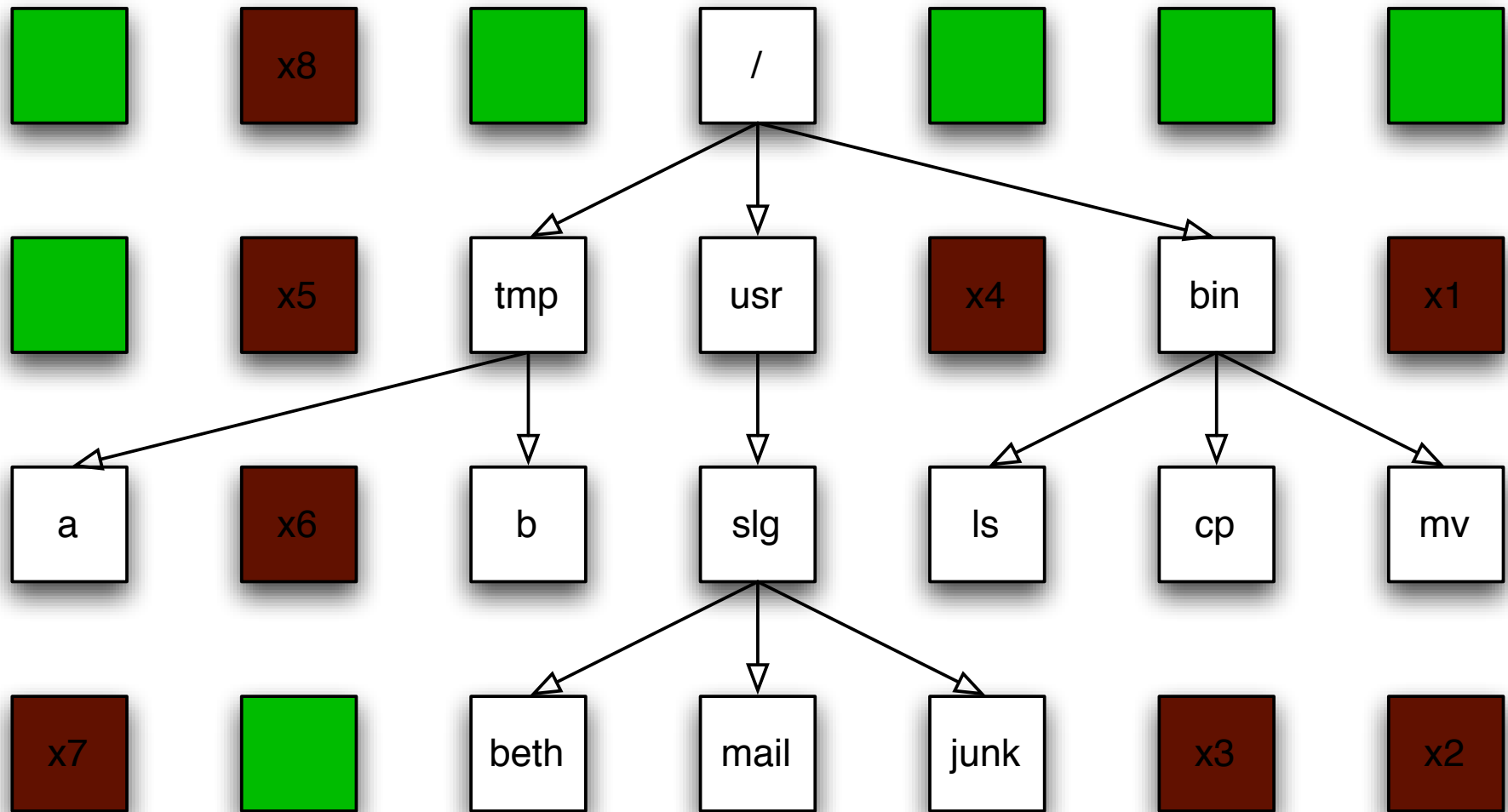
**Data on a hard drive is arranged in sectors.**



**The brown sectors indicate files that were deleted.**

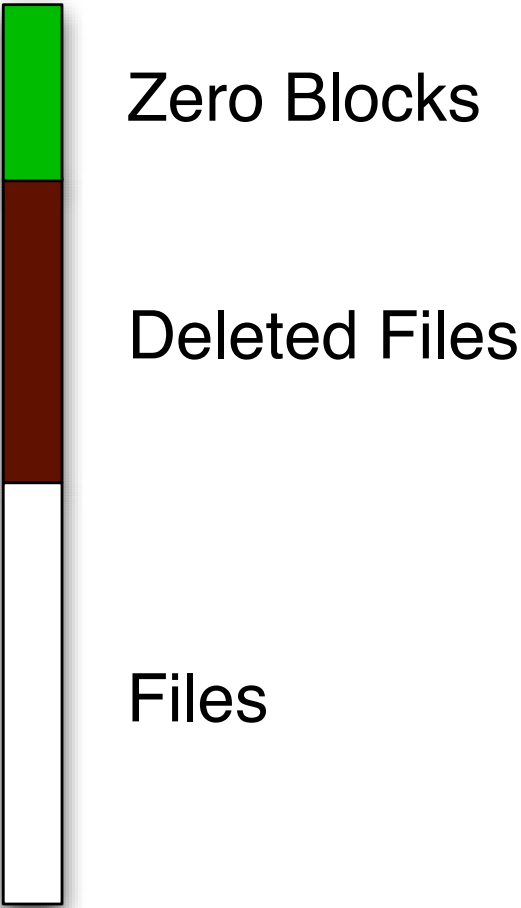
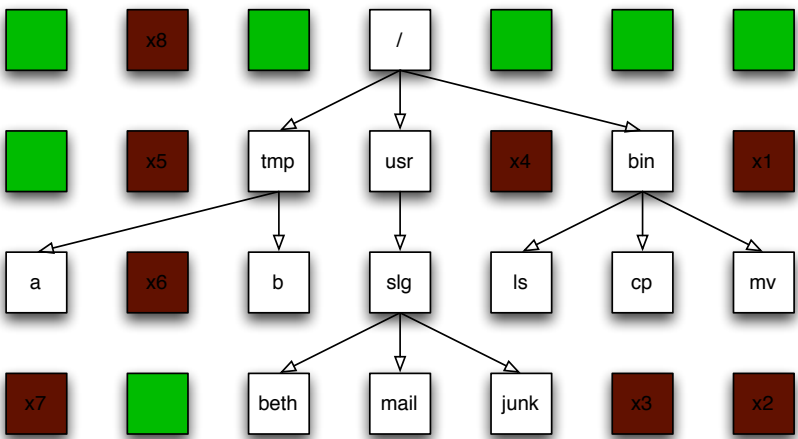


**Data on a hard drive is arranged in sectors.**

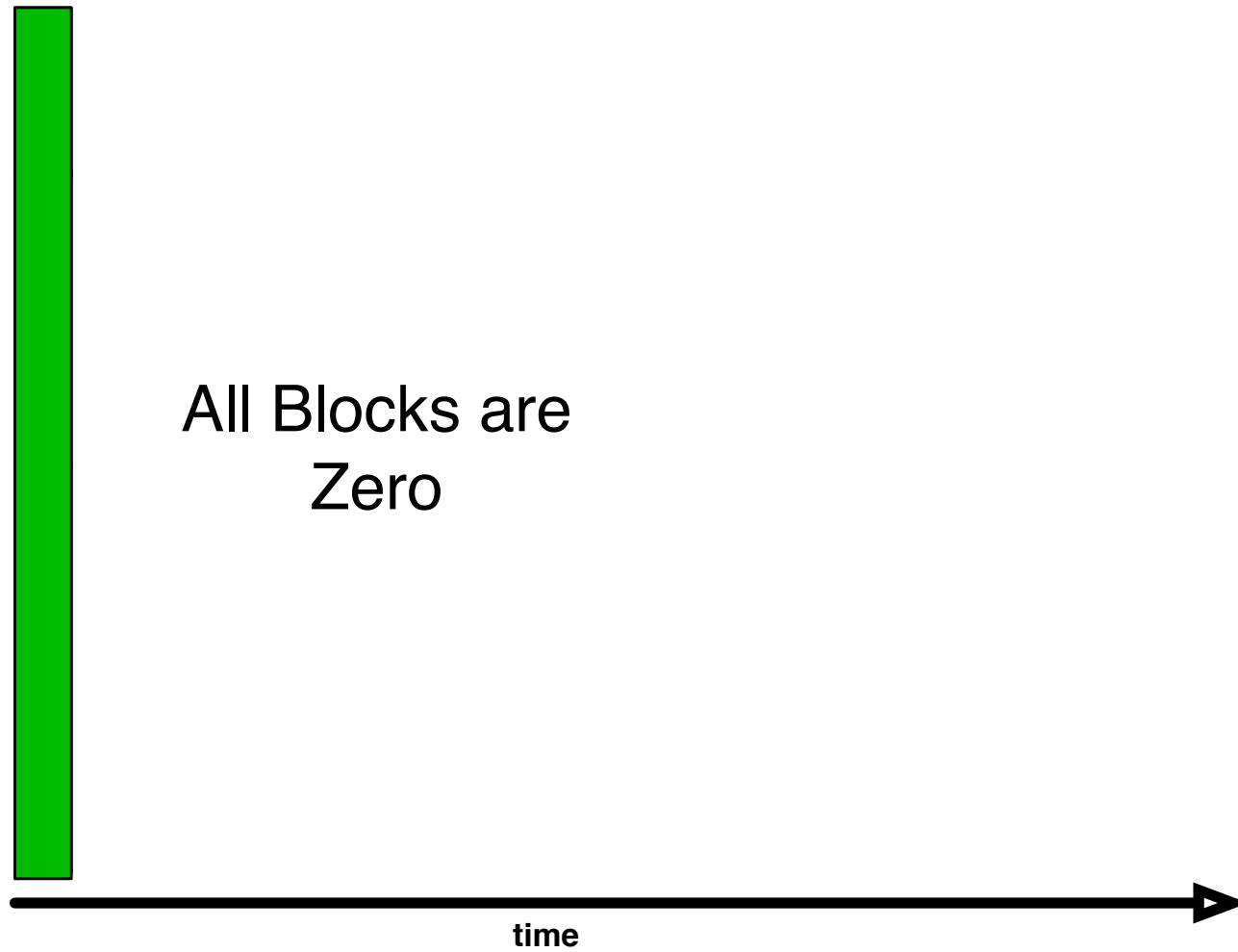


**The green sectors indicate sectors that were never used (or that were wiped clean).**

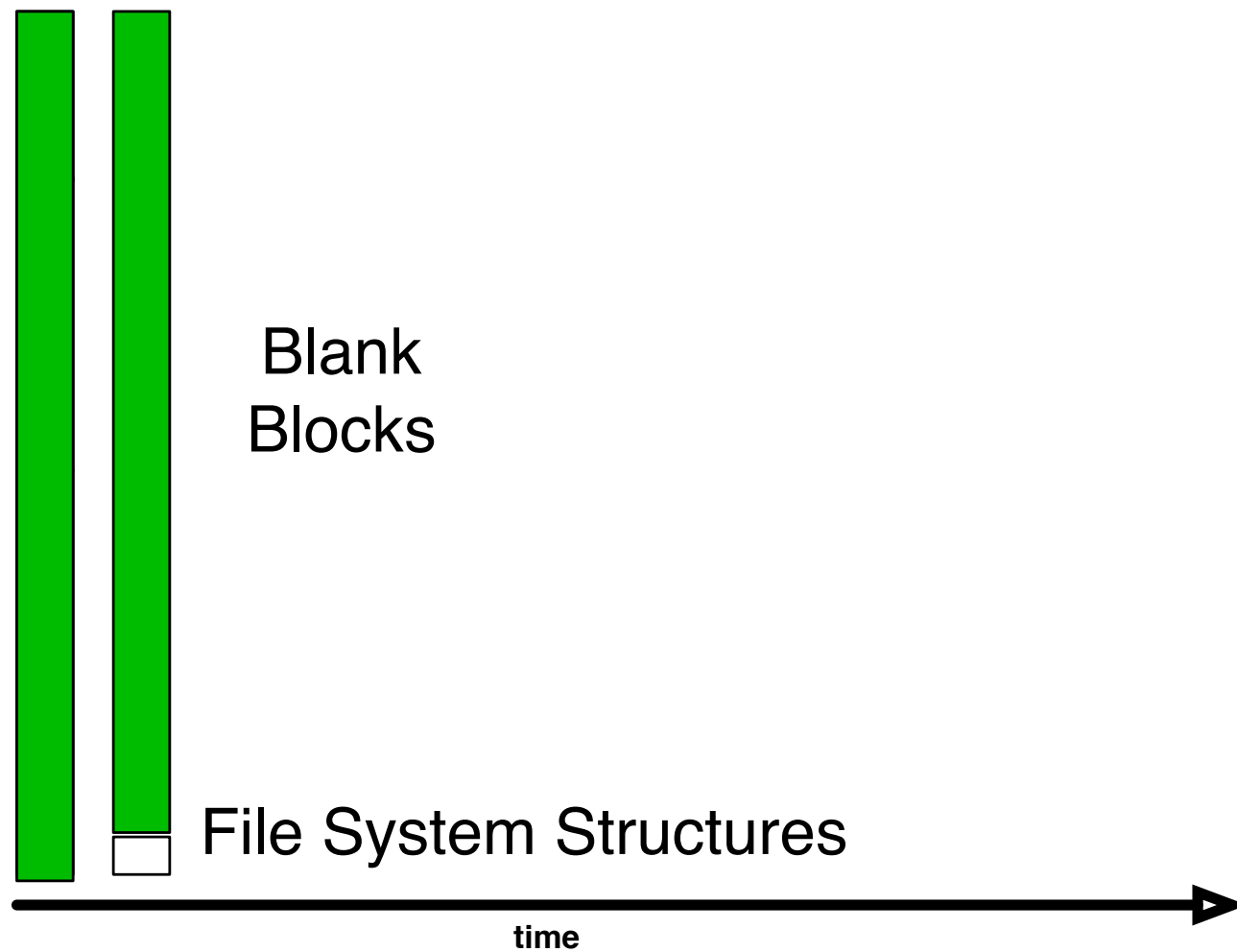
# Stack the disk sectors:



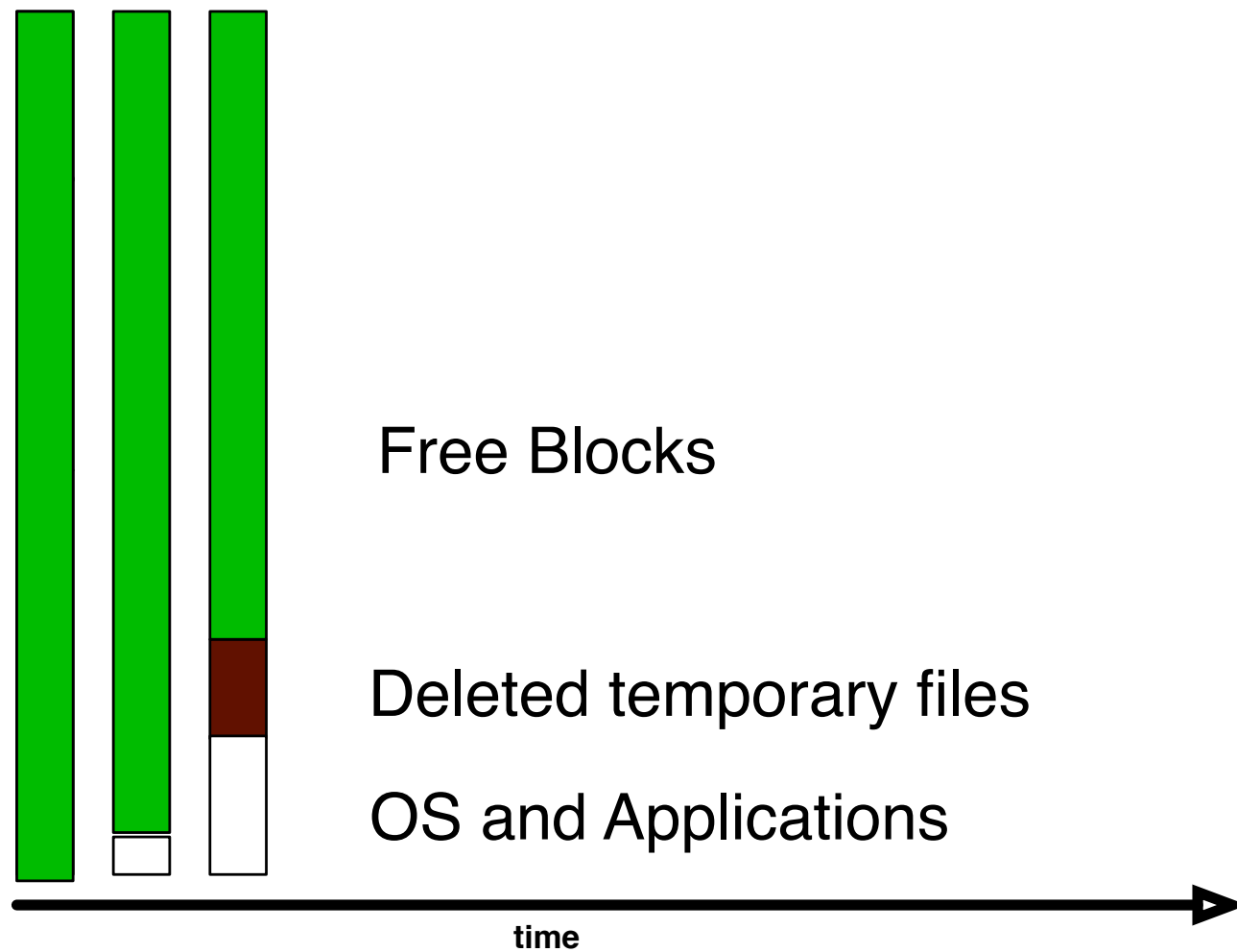
**NO DATA: The disk is factory fresh.**



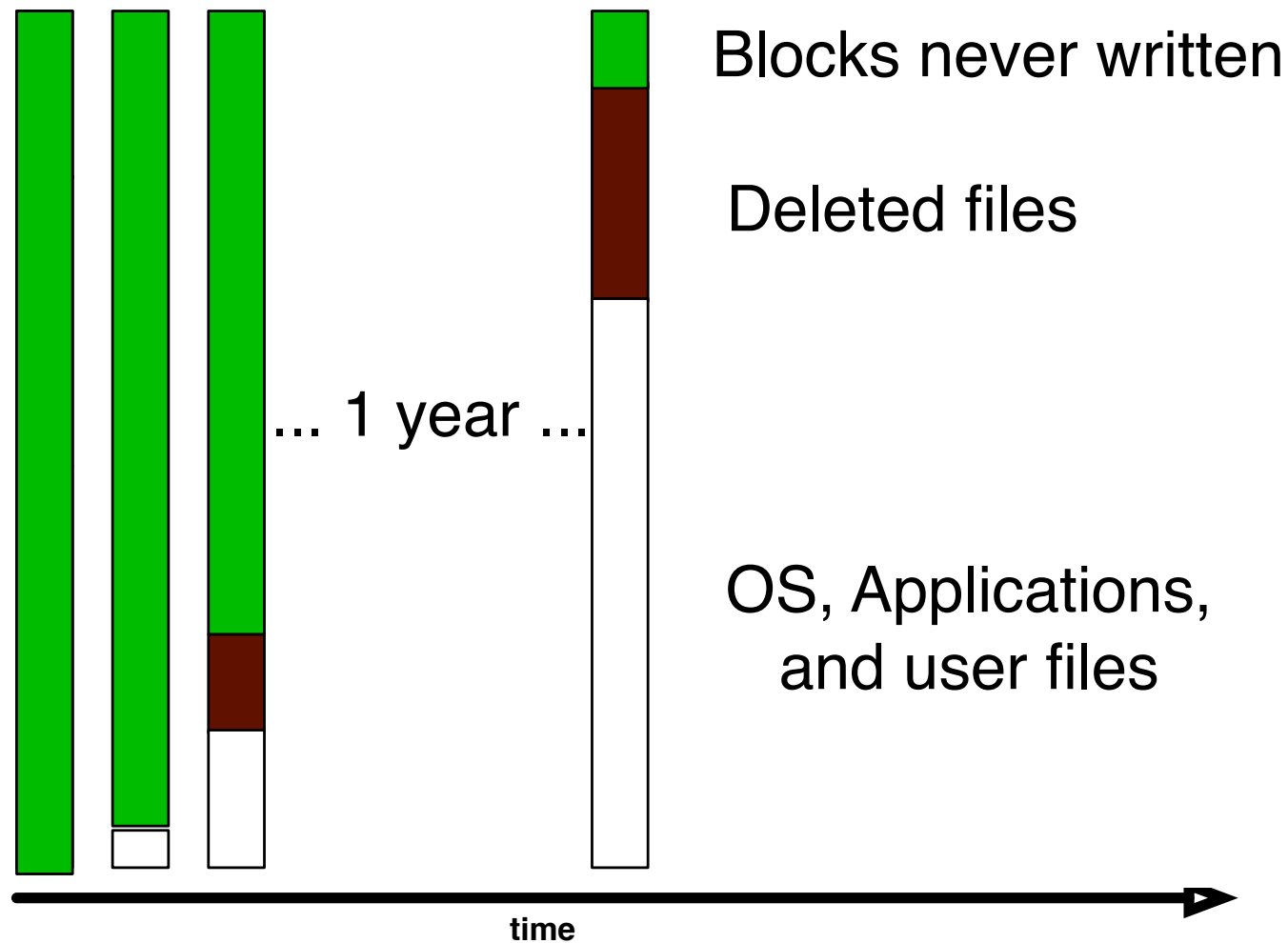
## FORMATTED: The disk has an empty file system



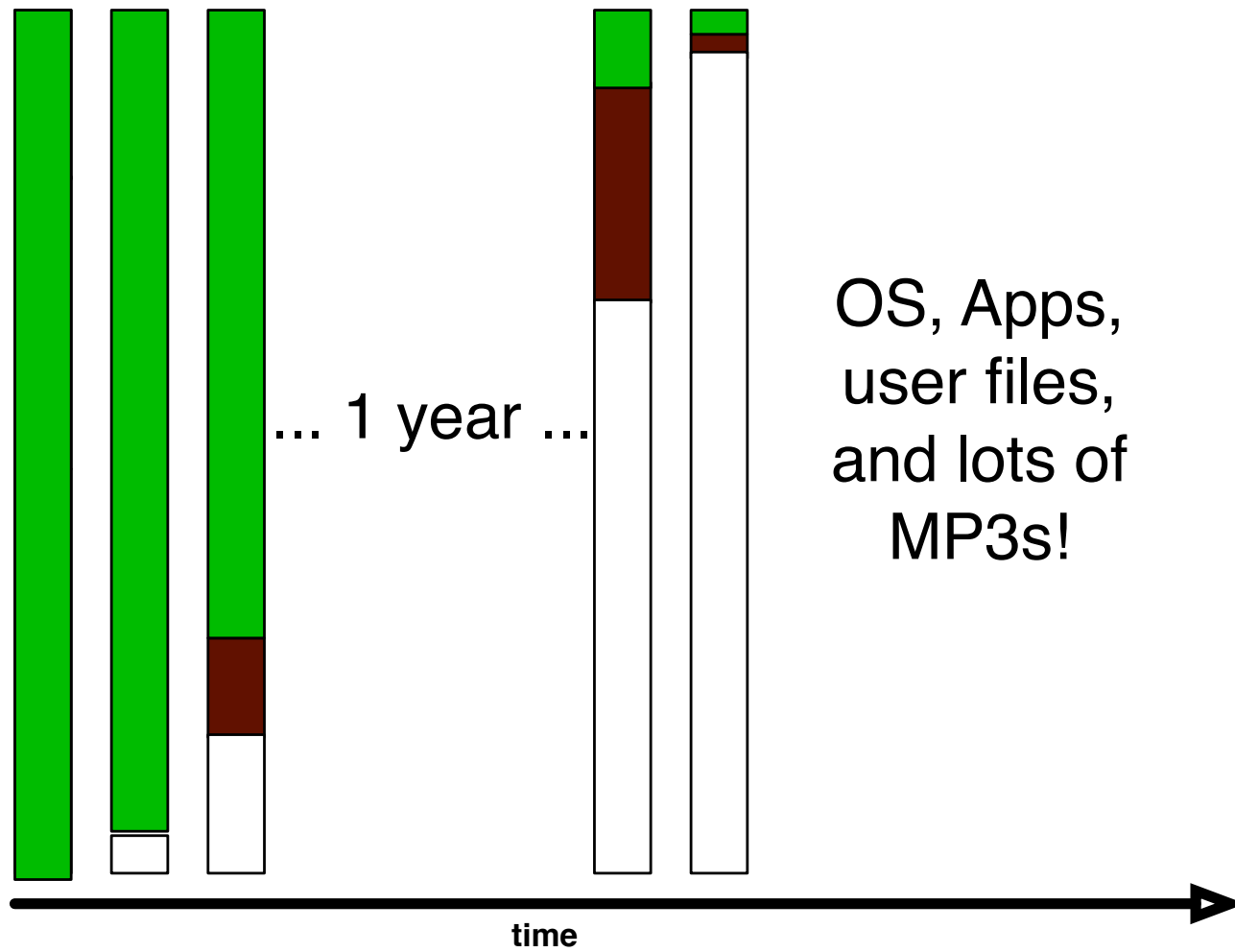
## AFTER OS INSTALL: Temp. files have been deleted



## AFTER A YEAR OF SERVICE

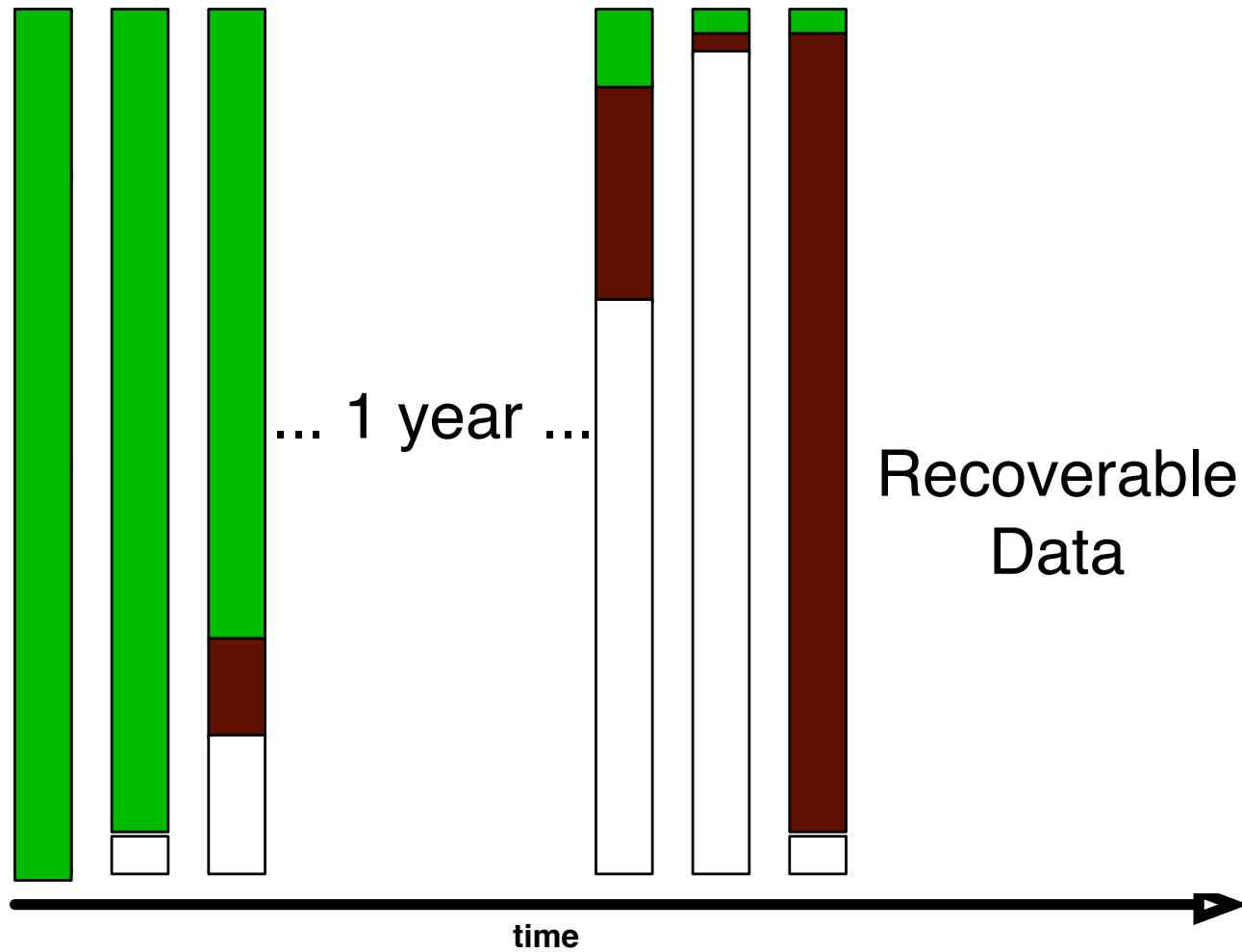


# DISK NEARLY FULL!

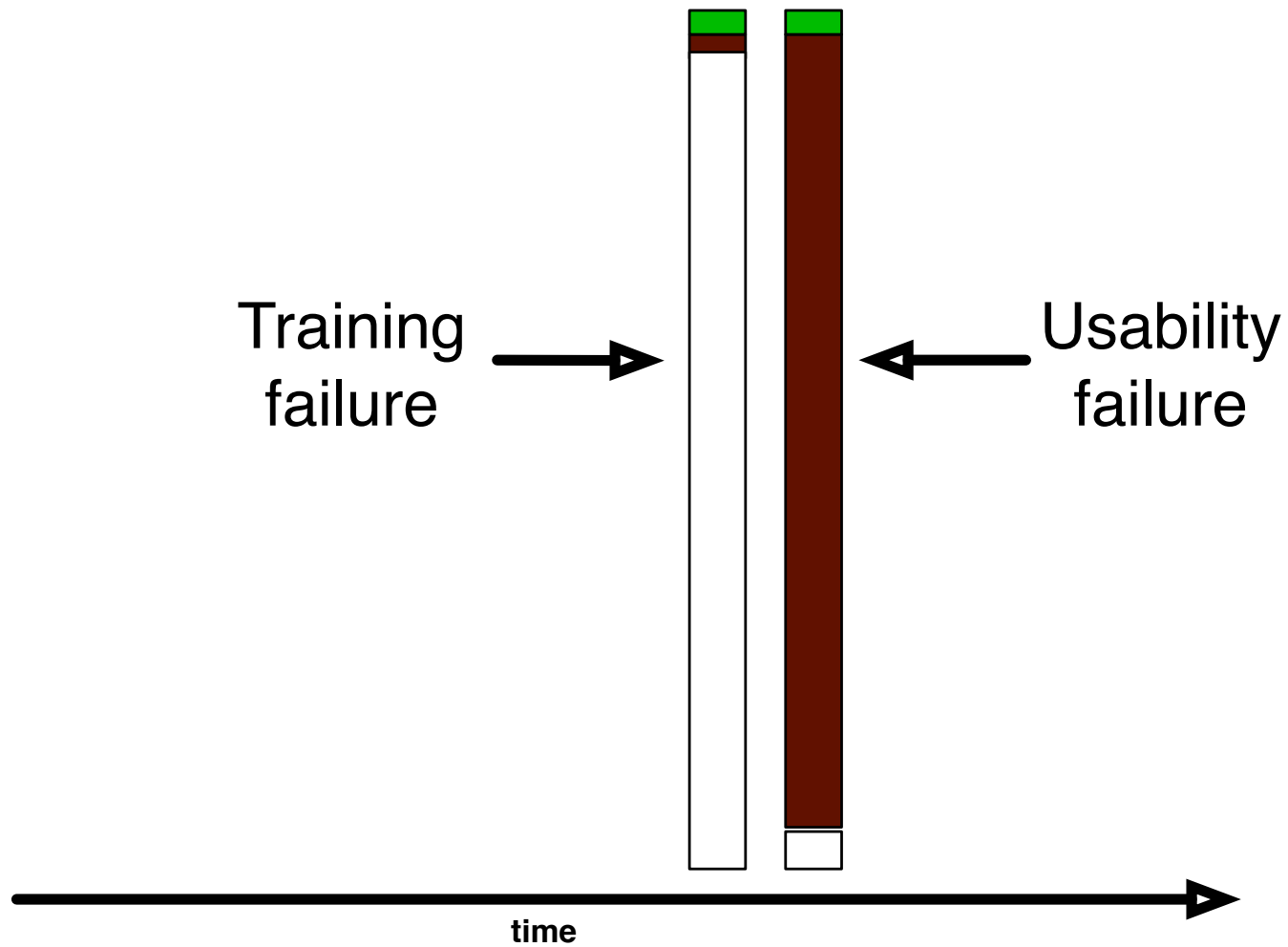




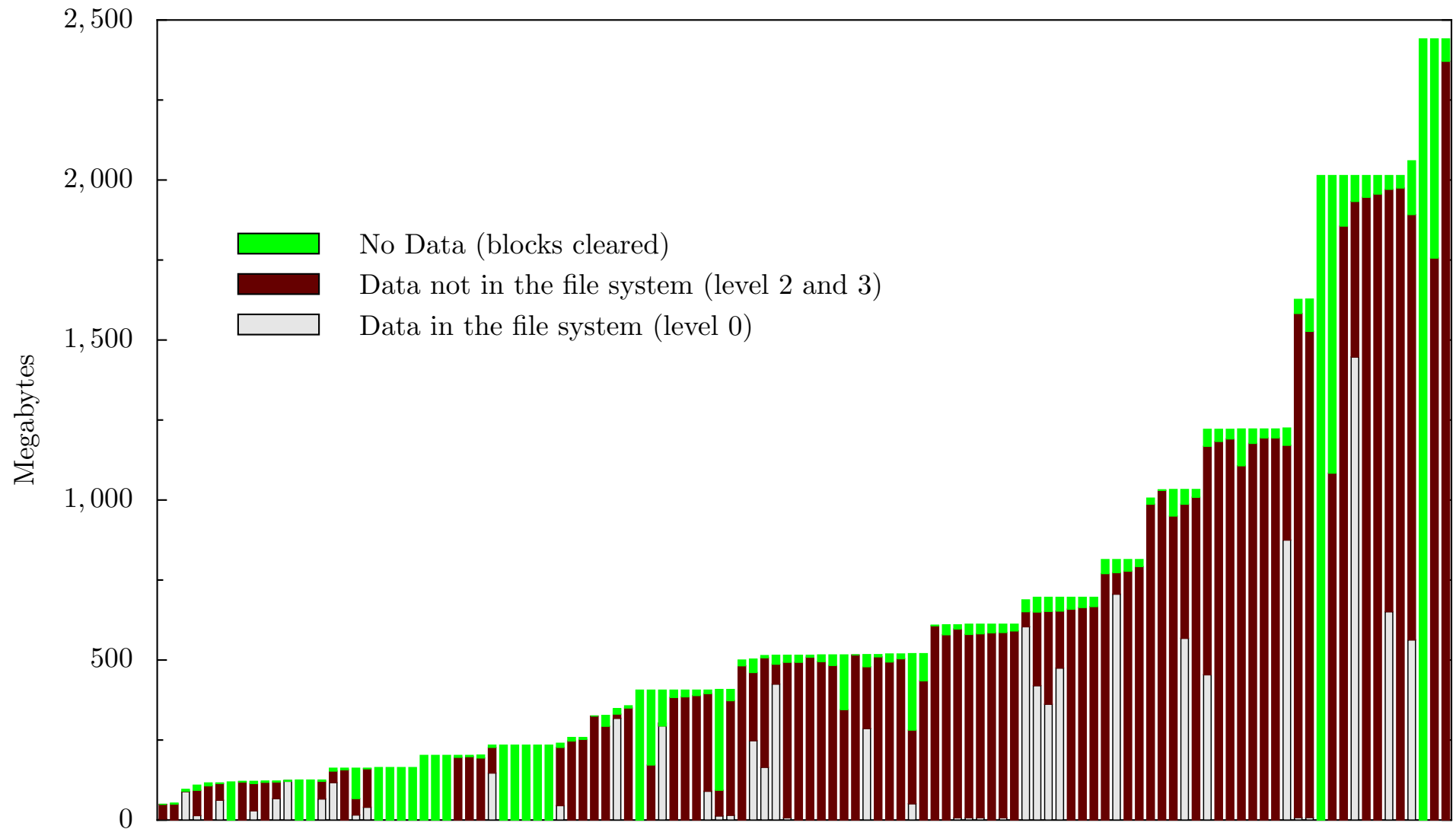
## FORMAT C:\ (to sell the computer.)



## We can use forensics to reconstruct motivations:



**Drives 1–236 are dominated by failed sanitization attempts.**



**..but training failures are also important.**

## Overall numbers

Drives Acquired:	236
Drives DOA:	60
Drives Images:	176
Drives Zeroed:	11
Drives "Clean Formatted:"	22
Total files:	168,459
Total data:	125G

## **Only 33 out of 176 working drives were properly cleared!**

- 1 from Driveguys — but 2 others had lots of data.
- 18 from pcjunkyard — but 7 others had data.
- 1 from a VA reseller — 1 DOA; 3 dirty formats.
- 1 from an unknown source — 1 DOA, 1 dirty format.
- 1 from Mr. M. who sold his 2GB drive on eBay.

## MD5 hashing allows the identification of files.

Interestingly, few unique files that had not been deleted:

File type	Unique Files
Microsoft Word files:	783
Microsoft Excel files:	184
Microsoft PowerPoint files:	30
Outlook PST files:	11
audio files:	977

**Conclusion: *most users* DELETED their files before discarding their drives.**

**But what *really* happened?**



**I needed to contact the original drive owners.**



# The *Remembrance of Data Passed Traceback Study.*

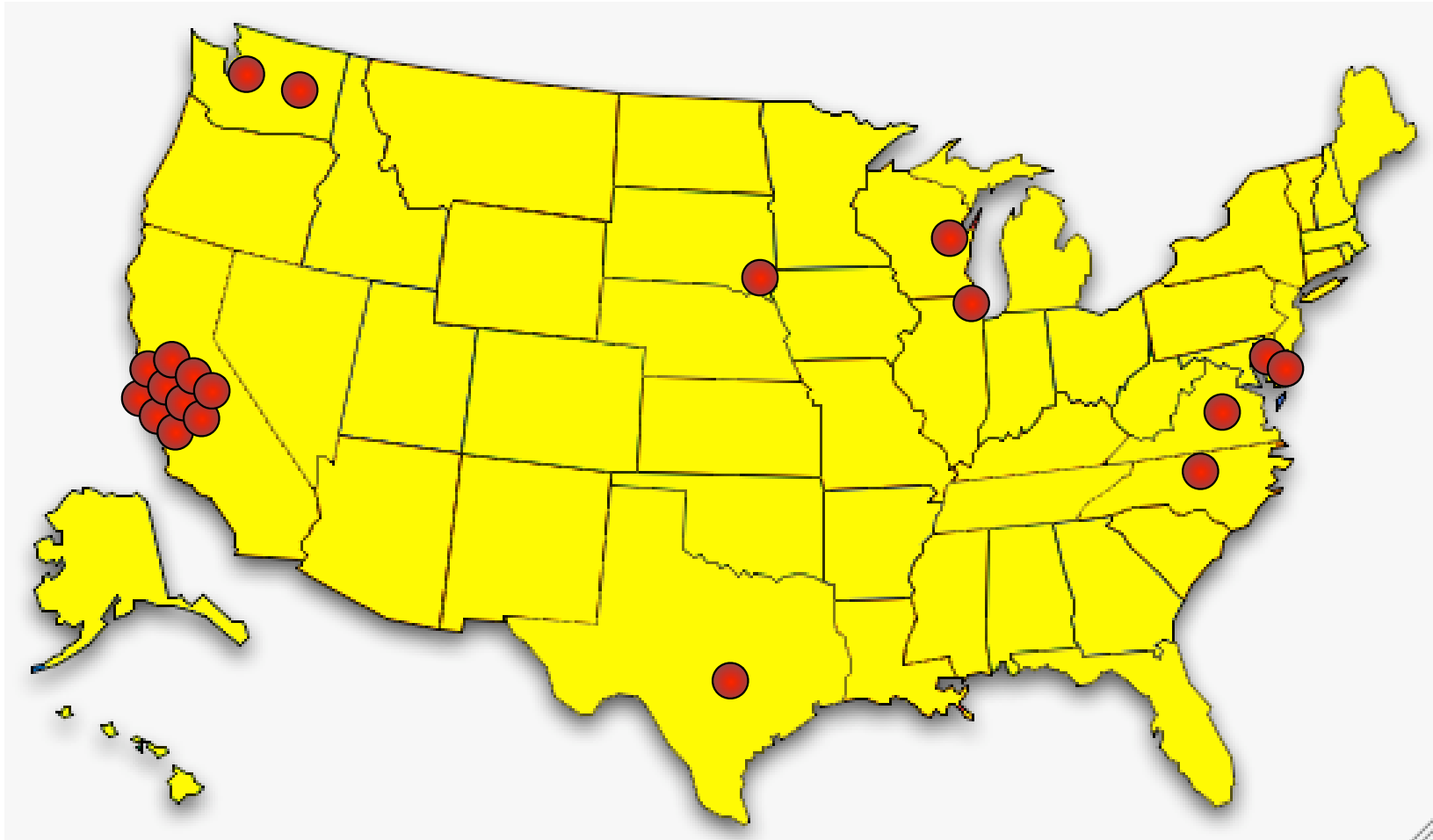
## [Garfinkel 05]

1. Find data on hard drive
2. Determine the owner
3. Get contact information for organization
4. Find the right person *inside* the organization
5. Set up interviews
6. Follow guidelines for human subjects work

```
06/19/1999  /:dir216/Four H Resume.doc
03/31/1999  /:dir216/U.M. Markets & Society.doc
08/27/1999  /:dir270/Resume-Deb.doc
03/31/1999  /:dir270/Deb-Marymount Letter.doc
03/31/1999  /:dir270/Links App. Ltr..doc
08/27/1999  /:dir270/Resume=Marymount U..doc
03/31/1999  /:dir270/NCR App. Ltr..doc
03/31/1999  /:dir270/Admissions counselor, NCR.doc
08/27/1999  /:dir270/Resume, Deb.doc
03/31/1999  /:dir270/UMUC App. Ltr..doc
03/31/1999  /:dir270/Ed. Coordinator Ltr..doc
03/31/1999  /:dir270/American College ...doc
04/01/1999  /:dir270/Am. U. Admin. Dir..doc
04/05/1999  /:dir270/IR Unknown Lab.doc
04/06/1999  /:dir270/Admit Slip for Modernism.doc
04/07/1999  /:dir270/Your Honor.doc
```

**This was a lot harder than I thought it would be.**

**Ultimately, I contacted 20 organizations between April 2003 and April 2005.**



## **The leading cause: betrayed trust.**

### Trust Failure: 5 cases

- ✓ Home computer; woman's son took to "PC Recycle"
- ✓ Community college; no procedures in place
- ✓ Church in South Dakota; administrator "kind of crazy"
- ✓ Auto dealership; consultant sold drives he "upgraded"
- ✓ Home computer, financial records; same consultant

**This specific failure wasn't considered in [GS 03];  
it was the most common failure.**

## **Second leading cause: Poor training and supervision**

Trust Failure: 5 cases

Lack of Training: 3 cases

- ✓ California electronic manufacturer
- ✓ Supermarket credit-card processing terminal
- ✓ ATM machine from a Chicago bank

**Alignment between the interface and the underlying representation would overcome this problem.**

## **Sometimes the data custodians just don't care.**

Trust Failure: 5 cases

Lack of Training: 3 cases

Lack of Concern: 2 cases

- ✓ Bankrupt Internet software developer
- ✓ Layoffs at a computer magazine

**Regulation on resellers might have prevented these cases.**



## **In seven cases, no cause could be determined.**

Trust Failure: 5 cases

Lack of Training: 3 cases

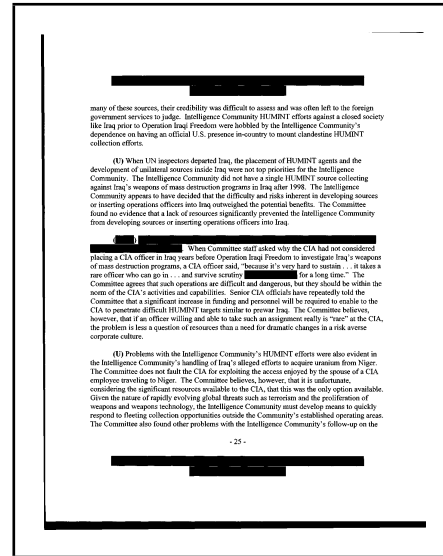
Lack of Concern: 2 cases

### **Unknown Reason: 7 cases**

- ✗ Bankrupt biotech startup
- ✗ Another major electronics manufacturer
- ✗ Primary school principal's office
- ✗ Mail order pharmacy
- ✗ Major telecommunications provider
- ✗ Minnesota food company
- ✗ State Corporation Commission

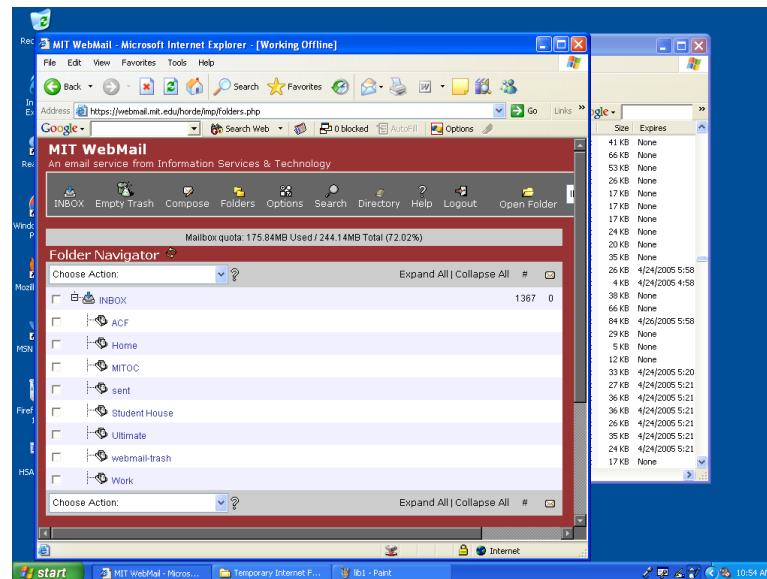
**Regulation might have helped here, too.**

# “Deleted” data can be recovered in other areas



## Document Files

## Web Browsers

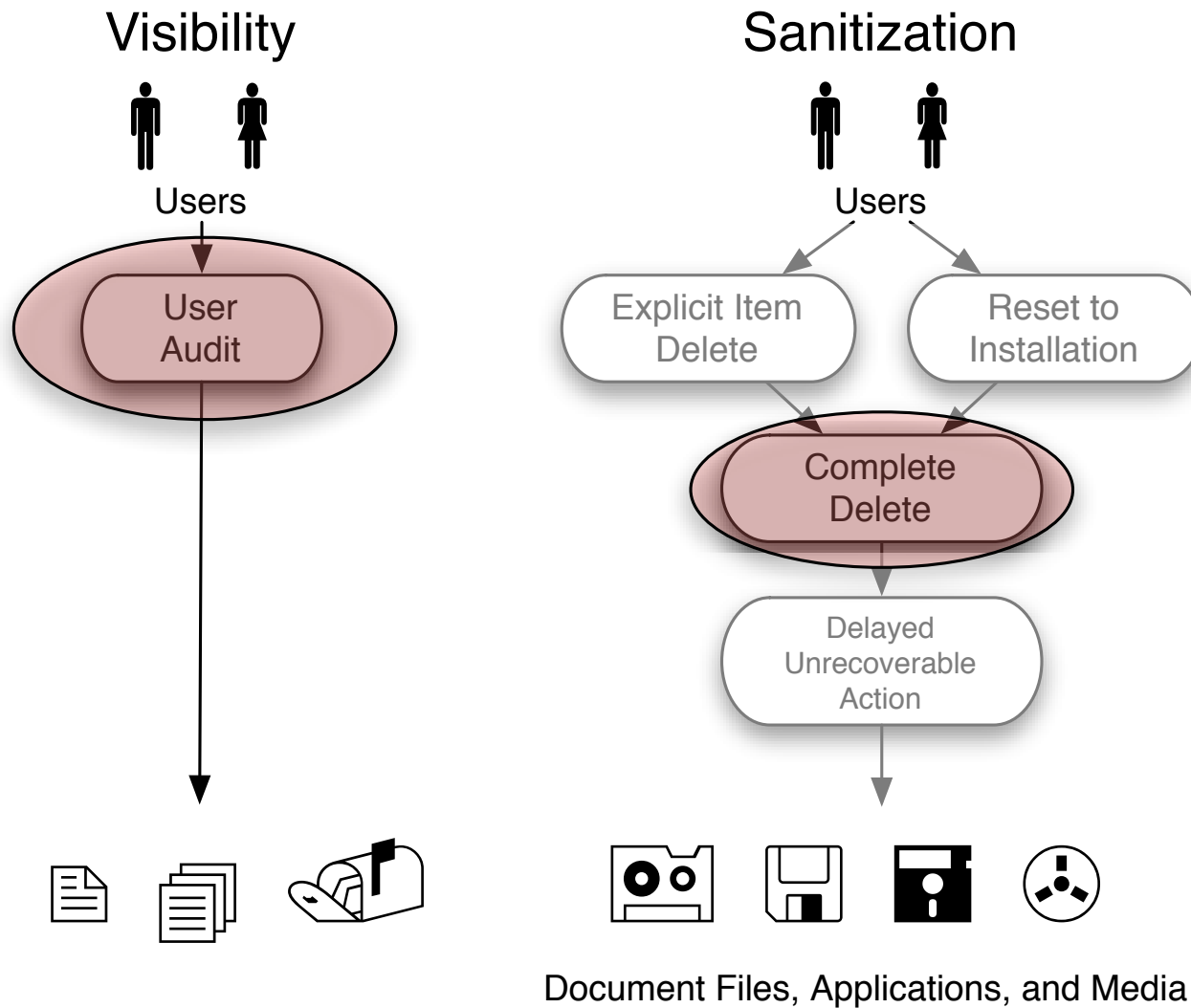


## Information is left in document files.

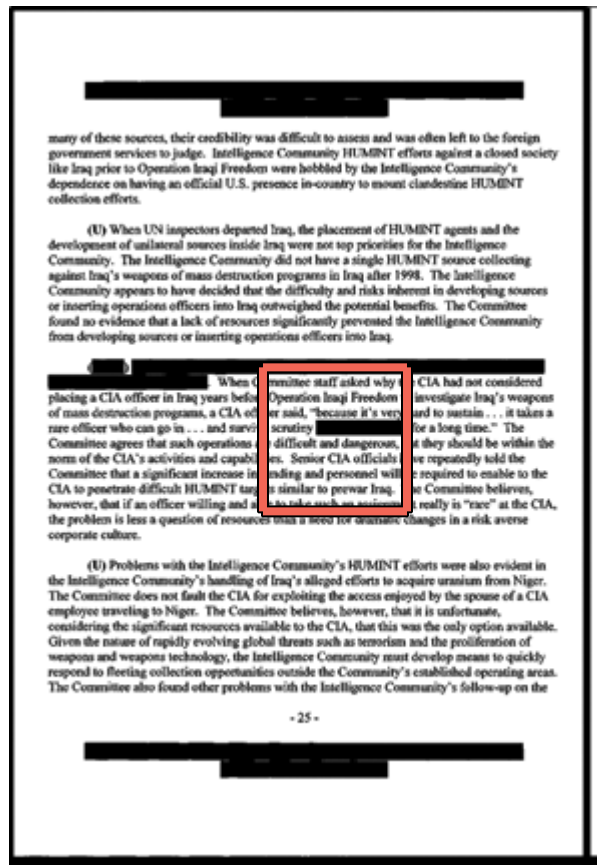
- The *New York Times* published a **PDF file** containing the names of Iranians who helped with the 1953 coup. [Young 00]
- US DoJ published a **PDF file** “diversity report” containing embarrassing redacted information. [Poulsen 03]
- SCO gave a **Microsoft Word file** to journalists that revealed its Linux legal strategy. [Shankland 04]
- Multinational Force-Iraq report

E. (U) Unit Experience in the Baghdad Area of Responsibility .....	8
1. (U) [REDACTED] Division .....	8
2. (U) [REDACTED] Brigade, [REDACTED] Division .....	9
3. (U) [REDACTED] Battalion .....	9
4. (U) [REDACTED] Battalion .....	10
F. (U) Findings .....	10

# The information leaked because two patterns were not implemented.



The Senate Foreign Intelligence Committee prevented leakage by *scanning* its redacted report on pre-war Iraq intelligence failures to create the PDF that it distributed.

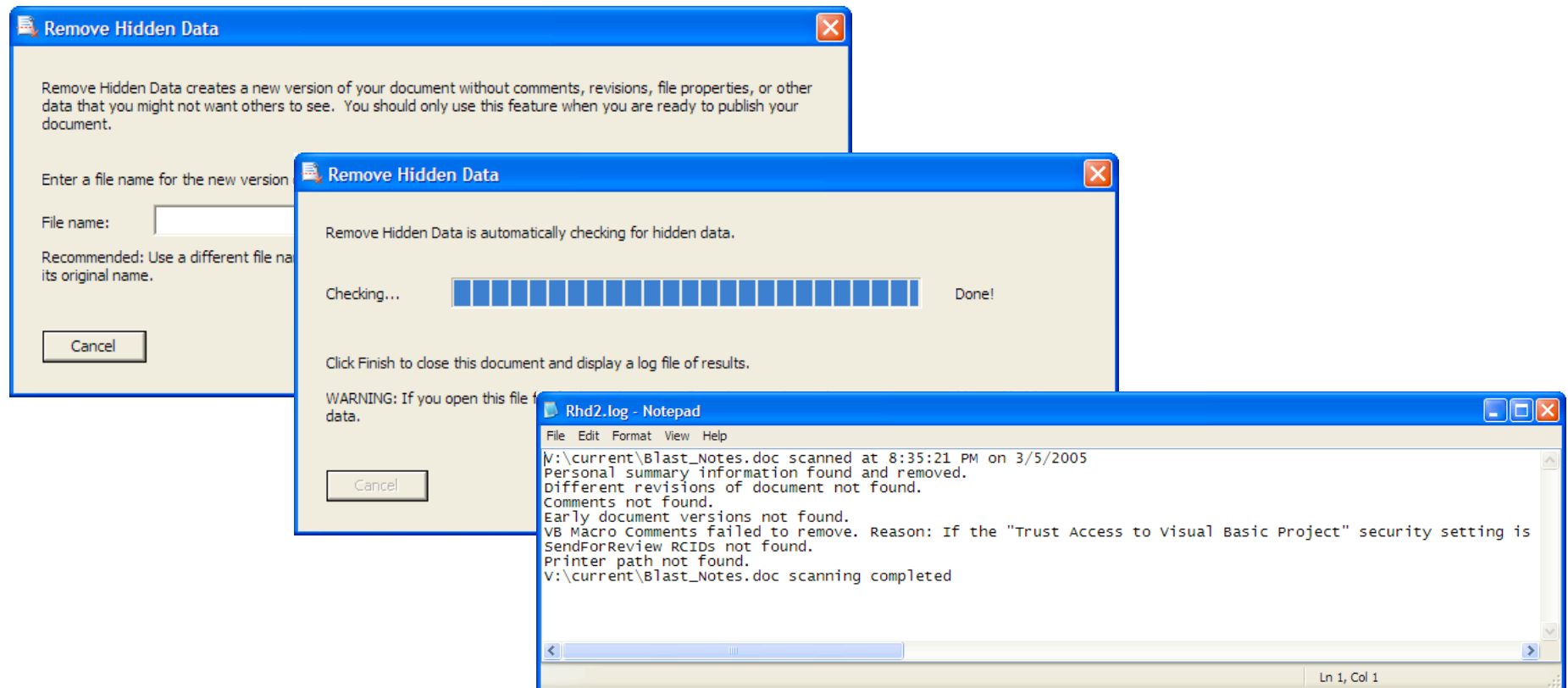


nittee staff asked w  
eration Iraqi Freed  
said, "because it's  
rutiny [redacted]  
difficult and danger  
Senior CIA offic  
ling and personnel  
similar to prewar Ir

This violates Section 503 (but they don't care).

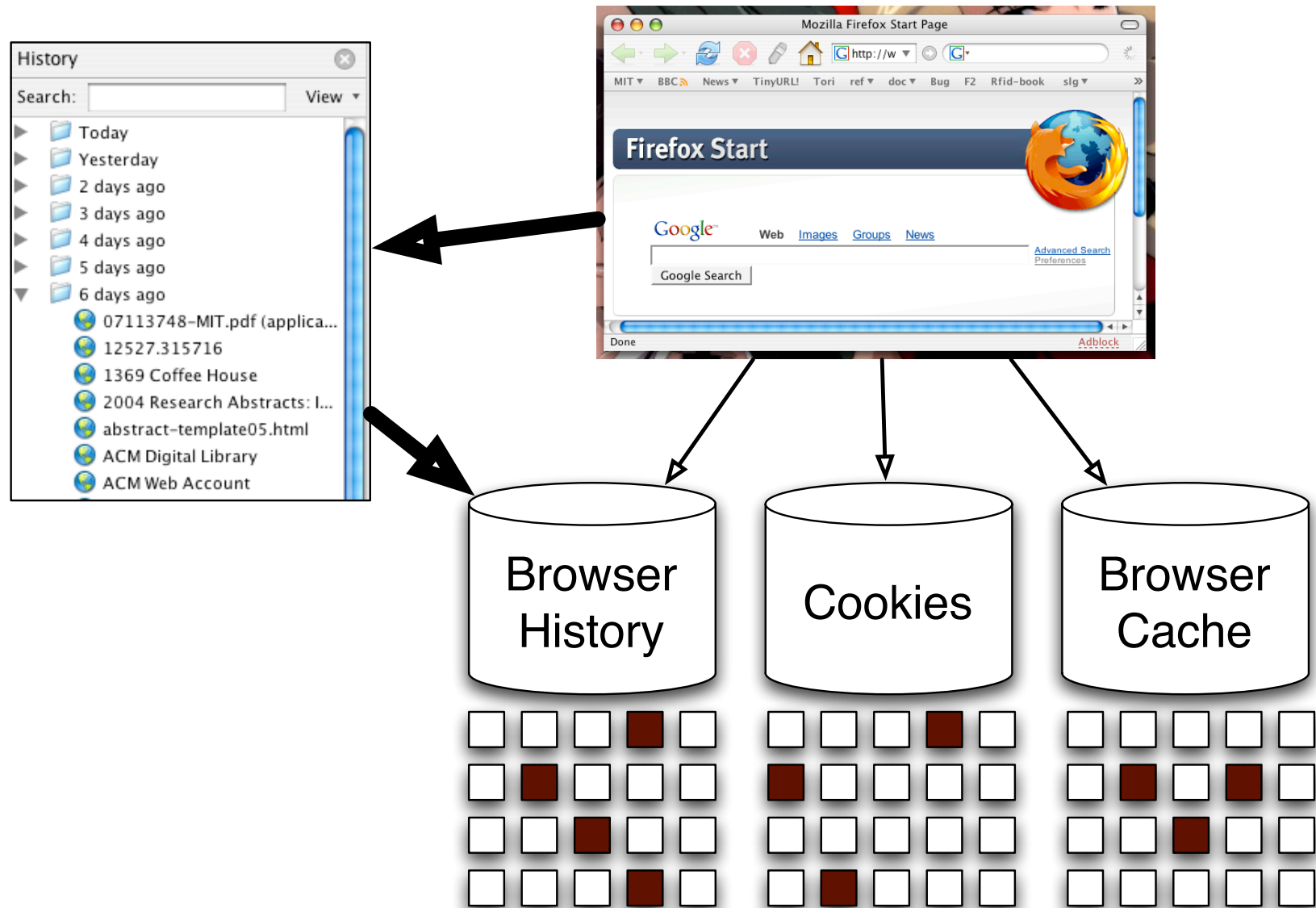


## Microsoft has tried to solve this problem with its “Remove Hidden Data” tool.



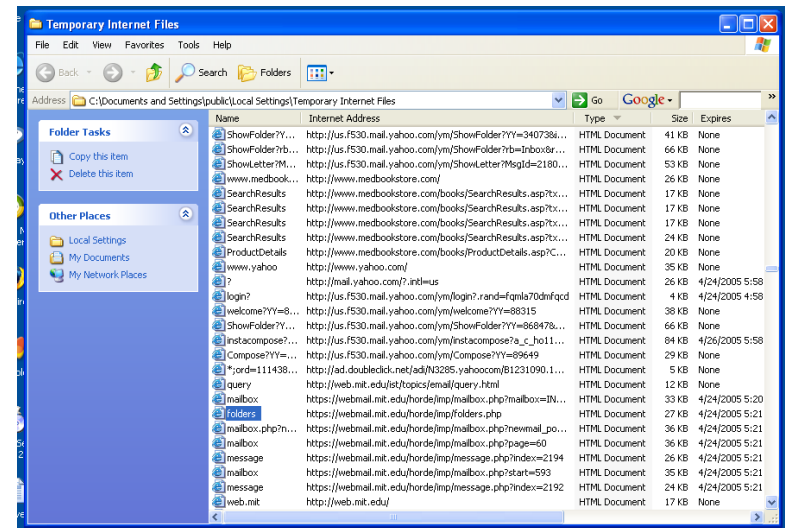
**RHD doesn't integrate into the flow of document preparation. The patterns-based analysis predicts that RHD will fail in many cases.**

# Information is left behind in web browsers.



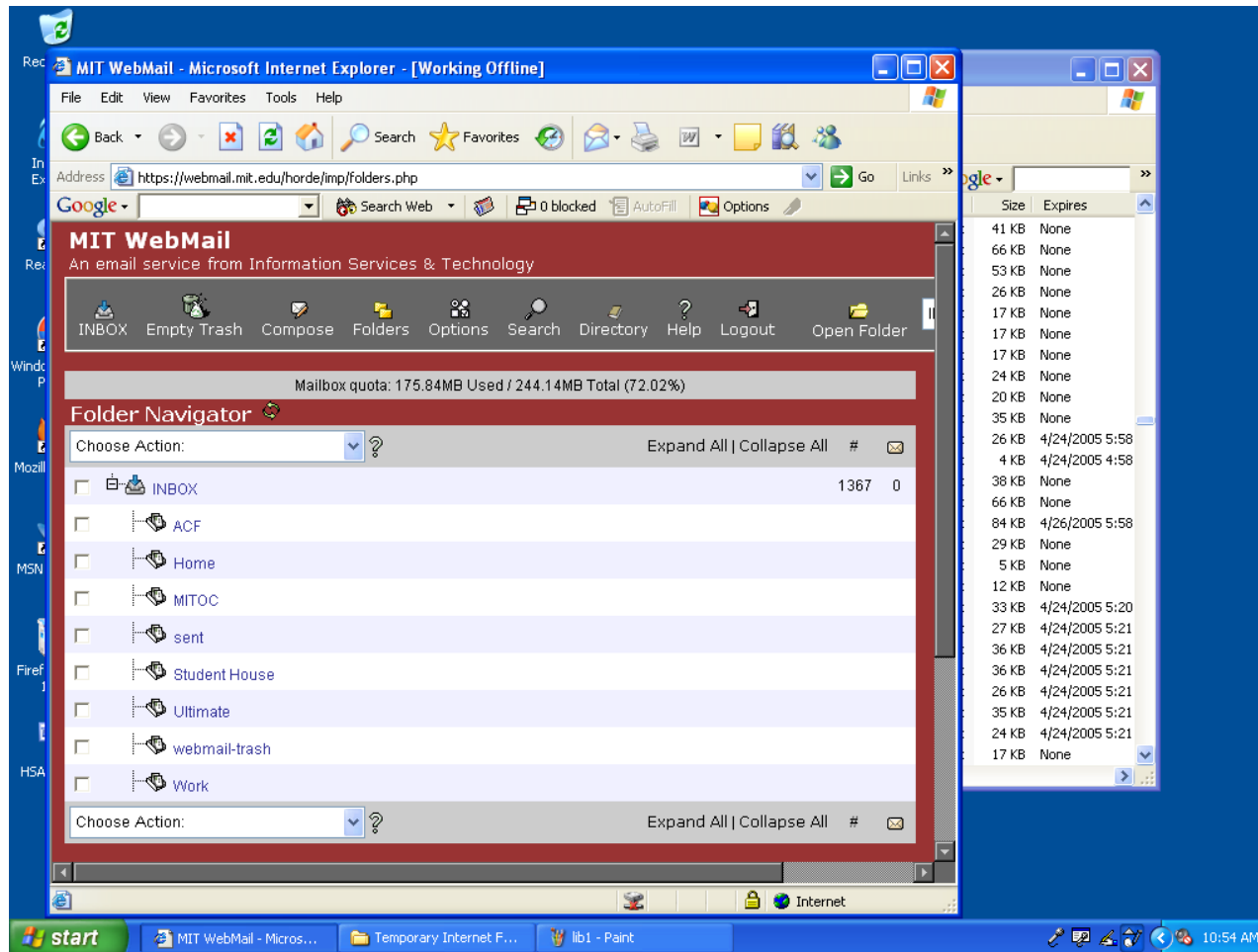
Two key problems: ① Deleted files; ② The cache

**In fact, a lot of information is left behind in web browsers.**



**MIT Humanities Library, April 25, 2005**

4 out of 4 computers had personal email in their browser caches.



The American Library Association recommends software that automatically purges caches on a *daily* basis. (It would be better to purge after each use.)

## Legislative reactions to this research:

### “Fair and Accurate Credit Transactions Act of 2003” (US)

- Introduced in July 2003.  
Signed December 2003.
- Regulations adopted in 2004, effective June 2005.
- Amends the FCRA to standardize consumer reports.
- Requires destruction of paper or electronic “consumer records.”

**Testimony:** <http://tinyurl.com/cd2my>

## Technical reactions to this research: “Secure Empty Trash” in MacOS 10.3.

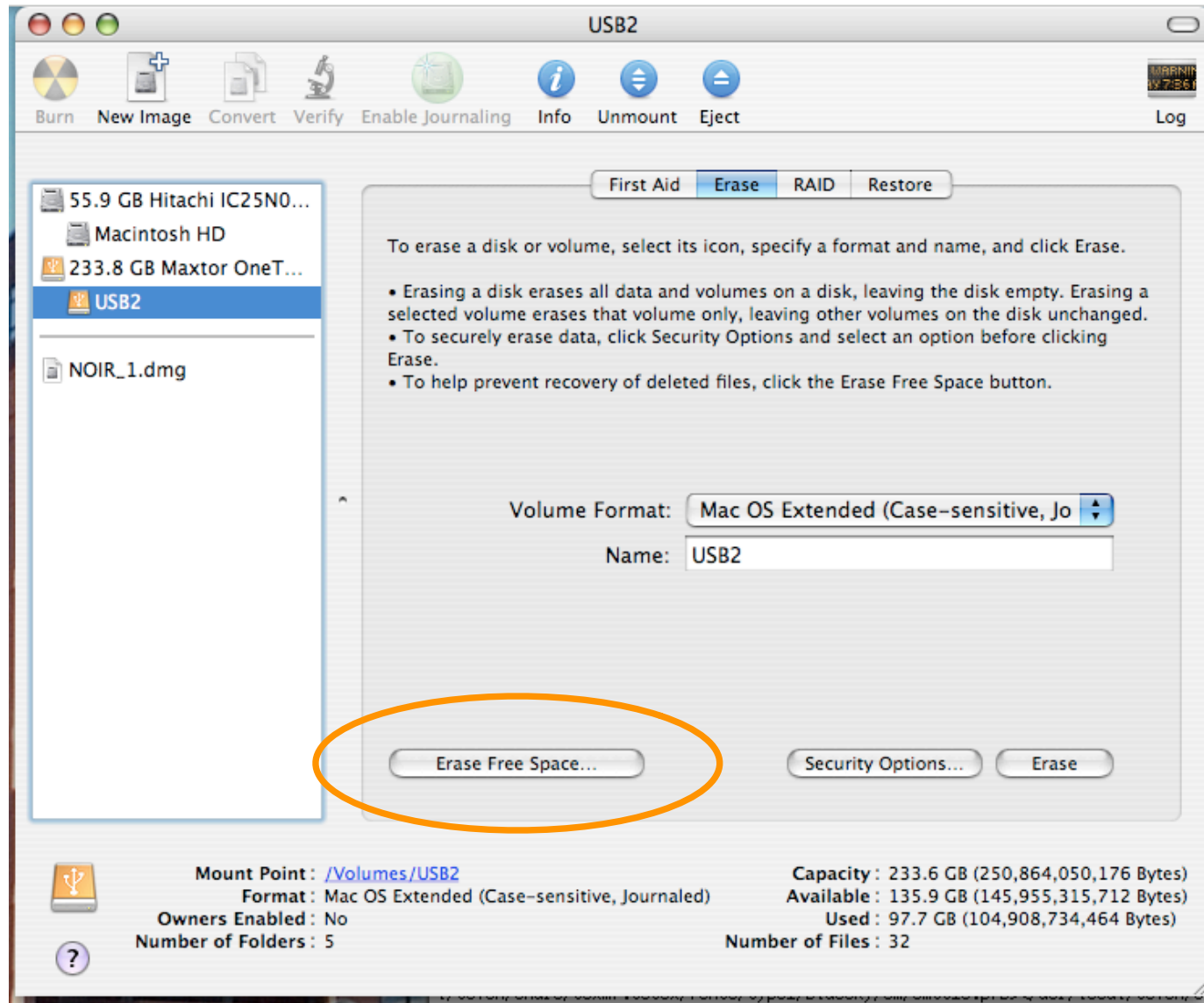


# Unfortunately, “Secure Empty Trash” is incomplete.

- Implemented in Finder (inconsistently)
- Locks trash can
- Can't change your mind

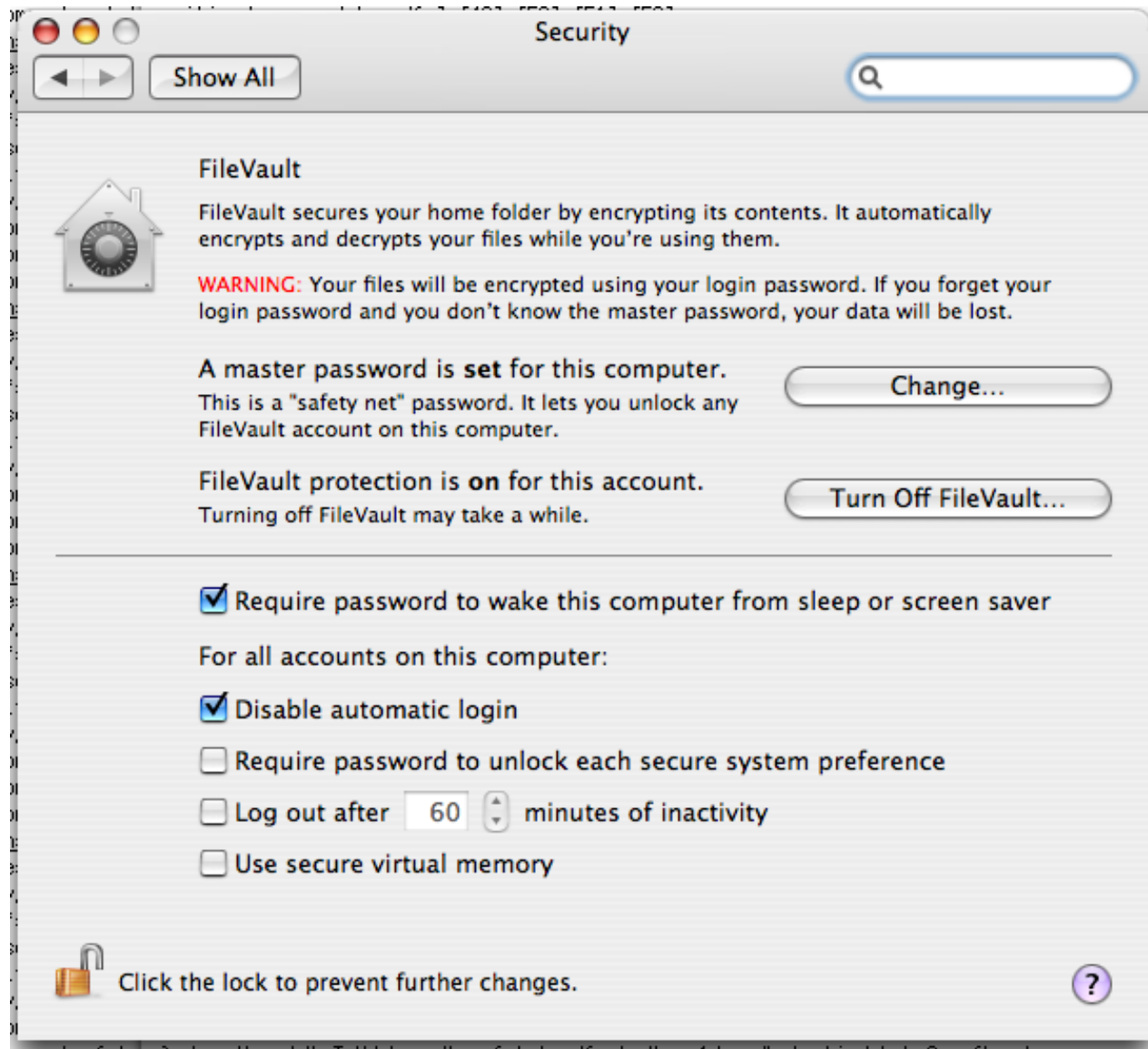


# MacOS 10.4 “Erase Free Space” makes a big file.





# MacOS “File Vault” gives users an encrypted file system.



# Future Work: Deploying Compete Delete

- Make FORMAT actually erase the disk.
- Make “Empty Trash” actually overwrite data.
- Integrate this functionality with web browsers, word processors, operating systems.
- Address usability dangers of clean delete.
- Analysis of “one big file” technique.

many of these sources, their credibility was difficult to assess and was often left to the foreign government services to judge. Intelligence Community HUMINT efforts against a closed society like Iraq prior to Operation Iraqi Freedom were hobbled by the Intelligence Community's dependence on having an official U.S. presence in-country to mount clandestine HUMINT collection efforts.

(U) When UN inspectors departed Iraq, the placement of HUMINT agents and the development of unilateral sources inside Iraq were not top priorities for the Intelligence Community. The Intelligence Community did not have a single HUMINT source collecting against Iraq's weapons of mass destruction programs in Iraq after 1998. The Intelligence Community appears to have decided that the difficulty and risks inherent in developing sources or inserting operations officers into Iraq outweighed the potential benefits. The Committee found no evidence that a lack of resources significantly prevented the Intelligence Community from developing sources or inserting operations officers into Iraq.

When Committee staff asked why the CIA had not considered placing a CIA officer in Iraq years before Operation Iraqi Freedom to investigate Iraq's weapons of mass destruction programs, a CIA officer said, “because it's very hard to sustain . . . it takes a rare officer who can go in . . . and survive scrutiny for a long time.” The Committee agrees that such operations are difficult and dangerous, but they should be within the norm of the CIA's activities and capabilities. Senior CIA officials have repeatedly told the Committee that a significant increase in funding and personnel will be required to enable the CIA to penetrate difficult HUMINT targets similar to prewar Iraq. The Committee believes, however, that if an officer willing and able to take such an assignment really is “rare” at the CIA, the problem is less a question of resources than a need for dramatic changes in a risk averse corporate culture.

(U) Problems with the Intelligence Community's HUMINT efforts were also evident in the Intelligence Community's handling of Iraq's alleged efforts to acquire uranium from Niger. The Committee does not fault the CIA for exploiting the access enjoyed by the spouse of a CIA employee traveling to Niger. The Committee believes, however, that it is unfortunate, considering the significant resources available to the CIA, that this was the only option available. Given the nature of rapidly evolving global threats such as terrorism and the proliferation of weapons and weapons technology, the Intelligence Community must develop means to quickly respond to fleeting collection opportunities outside the Community's established operating areas. The Committee also found other problems with the Intelligence Community's follow-up on the

- 25 -

## Questions?