# A Process for Verifiable Digital Evidence and Data Preservation

NSRL

National
Software
Reference
Library

***John Tebbutt and Doug White***

NIST United States Department of Commerce
National Institute of Standards and Technology

## Disclaimer

Trade names and company products are mentioned in the text or identified. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.

## Statement of Disclosure

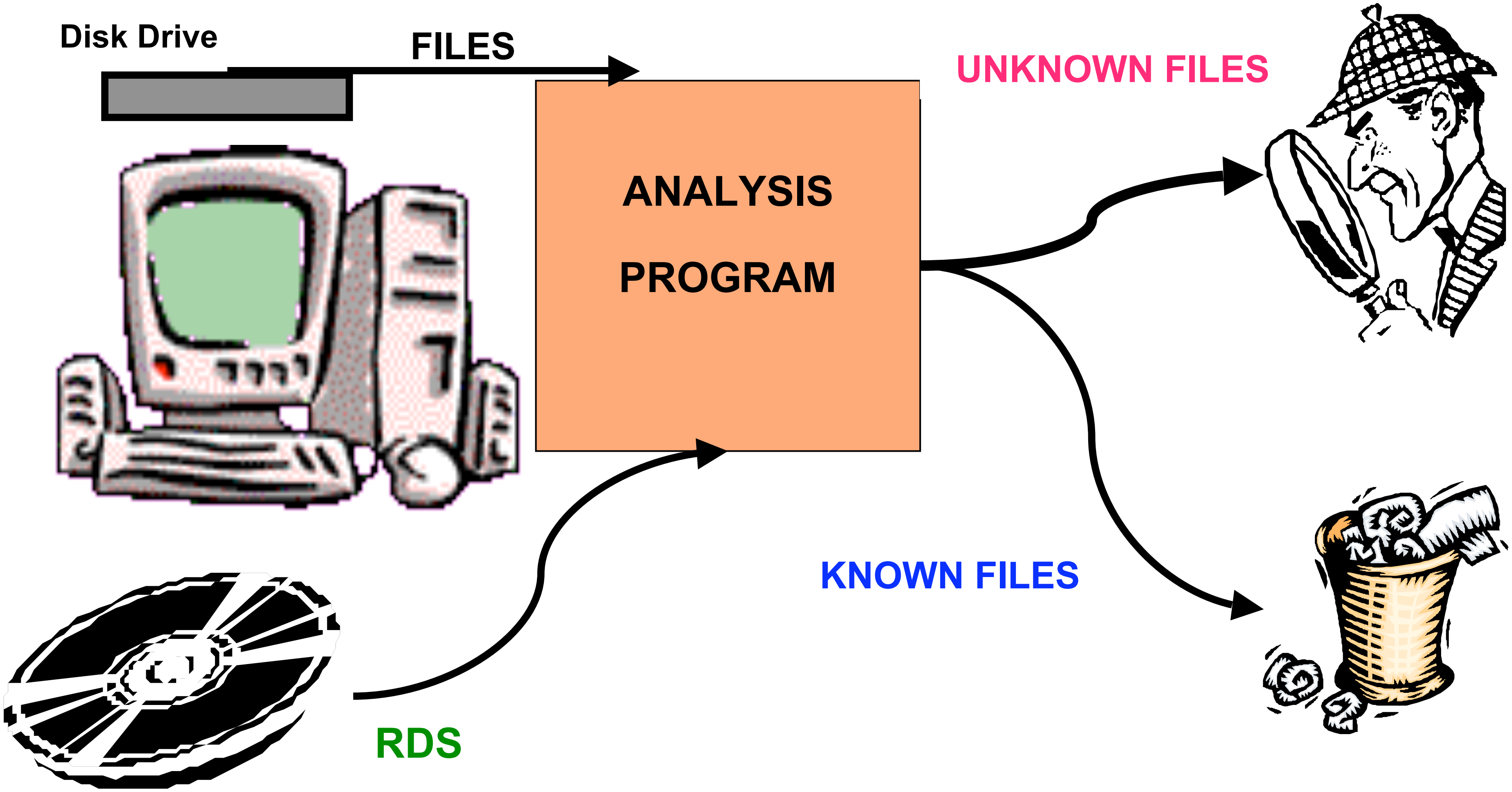
This research was funded by the National Institute of Standards and Technology Office of Law Enforcement Standards, the Department of Justice National Institute of Justice, the Federal Bureau of Investigation and the National Archives and Records Administration.

# What is the NSRL?

- **A physical collection of software**
- **A database of meta-information**
- **A published subset of the database, the Reference Data Set (RDS)**

The NSRL collects software from variou sources and provides file profiles compute from this software as a Reference Data Se (RDS) of file identifying information (FII).

# NSRL RDS: Data Reduction for Investigators

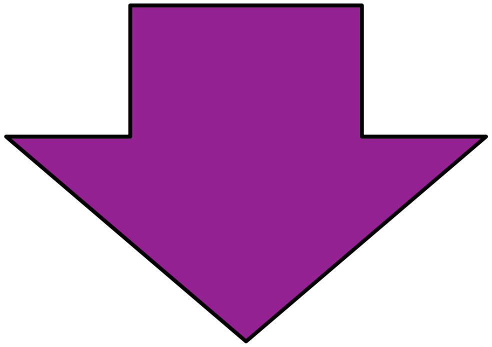


# What sort of Files?

- System files
  - Operating system executables
  - Hardware drivers
  - Static data structures (e.g. configuration files)...
- Application Files
  - Executables
  - Static data/configuration files...
- User-generated files
  - Documents, spreadsheets, etc; media; email; browser cache

# Evidentiary Value

- Trusted installation media
- Trusted processing environment
- Evidence locker storage



**TRACEABILITY**

# Trusted Installation Media

Shrink-wrapped physical media

- Minimize risk of tampering
- No downloads

Reputable source

- Manufacturer, large retailer, developer subscription
- Donations from trusted contacts (law enforcement)

Read-only media (CD, DVD)

- No floppies, CD-Rs, USB drives, etc.

# Trusted Processing Environment

Restricted access to systems and software

- No unauthorized physical access to systems

Isolated network

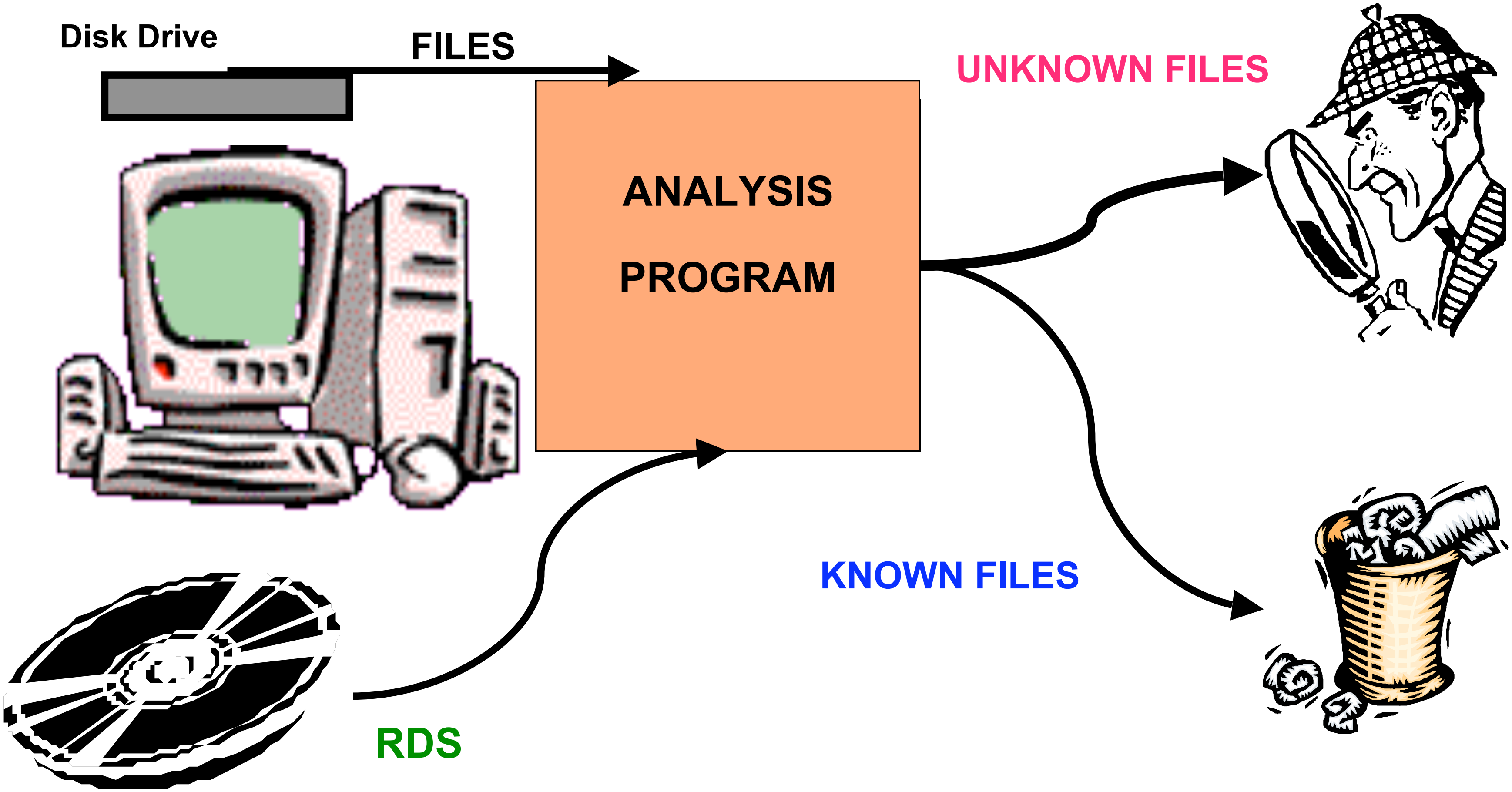
- No remote access to systems

Trusted in-house developed and tested software

# Evidence Locker Storage

- ✓ Restricted Access with Intrusion Detection
- ✓ Evidence Tracking System
- ✓ Periodic Auditing
- ✓ Minimal Human Contact

# NSRL RDS: Data Reduction for Investigators



# Data Preservation - NARA

- "...serves American democracy by safeguarding and preserving the records of our Government..."
- Yes, everything...

THE

NATIONAL ARCHIVES

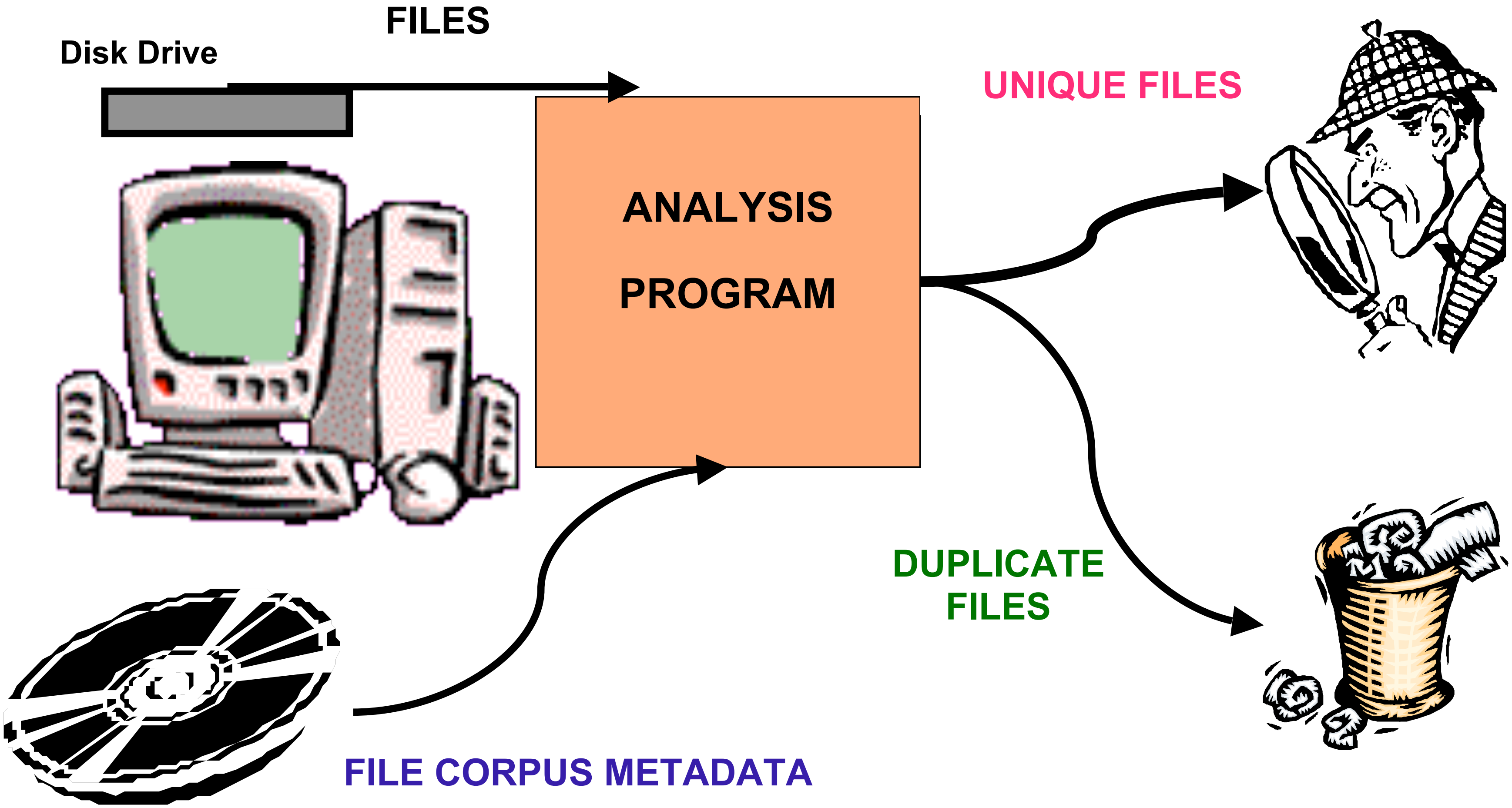
ARCHIVES.GOV

Democracy Starts Here.

# NARA: Need for automated data reduction

- Currently working on electronic records of the George H.W. Bush administration
- Estimated 9TB of data from current (George W Bush) administration (to August 2006)
- Minimal automation of analysis
- Much data is duplicated and may not be of interest, e.g. computer system files

# NARA: Data Reduction for Archivists



# NARA: NSRL Model

- Routinely process large numbers of files
  - generate and collect file metadata
- Application agnostic
  - NSRL: software installation media
  - NARA: computer hard drive data
- Easily extensible
  - additional metadata types

# NARA: Metadata types

- NSRL FII comprises “hard” metadata:
  - file name, file size, creation/modification times, etc
  - cryptographic hash of file contents
- Extend with “soft” metadata:
  - “fuzzy” hashes
    - identify similar files and quantify similarity
  - block hashes
    - cryptographic hashes of files' constituent data blocks

# NARA: Data reduction

- NSRL FII identifies:
  - unique files
    - ➔ preserve as “master” copy/copies
  - files with duplicate content
    - ➔ preserve only metadata which differ, e.g. file name, source computer, etc.
    - ➔ refer to “master” file for file contents
- No human input required
- Tests yielded 30% - 50% reduction

# NARA: Data management

- Simple automated file classification
- “Soft” metadata identify similar documents
  - ➔ Group for human analysis of file clustering
  - ➔ Analyst determines which files to retain
- Fast, foolproof – not intelligent

# Thank You

John Tebbutt
Computer Scientist
National Software Reference Library

National Institute of Standards and Technology
100 Bureau Drive STOP 8970
Gaithersburg, MD 20899-8970

# Contacts

**John Tebbutt**
**www.nsrl.nist.gov**
**nsrl@nist.gov**

**Barbara Guttman**
**Software Diagnostics & Conformance Testing Division**
**barbara.guttman@nist.gov**

**Sue Ballou, Office of Law Enforcement Standards**
**Rep. For State/Local Law Enforcement**
**susan.ballou@nist.gov**