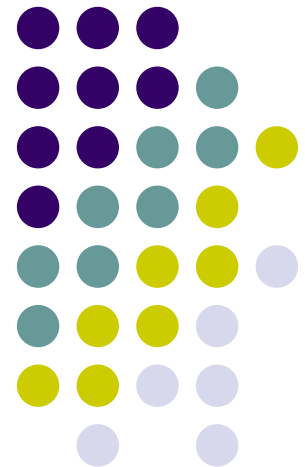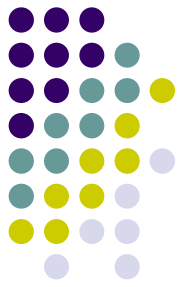# Gone With the Wind? Thoughts From the Archivist's Lawyer on Managing The Government's Ephemeral Electronic Records

RACO 2008 Atlanta
NARA's 5th Annual Regional Records Administration Conference
September 9, 2008

Jason R. Baron
Director of Litigation
Office of General Counsel
National Archives and Records Administration

Email vs The Phone

April 22, 2003

E-mail More Important Than the Phone In Business, Study Shows

By Antone Gonsalves

The reign of the telephone is over.

After more than 125 years since inventor Alexander Graham Bell yelled those historic words, "Come here Mr. Watson, I want to see you," the telephone is no longer the most valuable means of business communications. Its usurper: e-mail.

A survey of businesspeople at 387 organizations found that 80 percent believed e-mail was more important than the telephone in communicating with coworkers, customers, or partners . . . .

Just as surprising for analysts were findings that 74 percent of the respondents believed being without e-mail would present a greater hardship than losing telephone service.

EMPA – Remarks on House Passage

"With the advent and proliferation of electronic communication, the time has come for us to improve the methods for preserving records. Gone are the days when correspondence was sent solely using paper. E-mail has become the primary method of communication, and it is imperative that we develop procedures to stay current with the evolving technologies."

--The Hon. Carolyn B. Maloney, July 9, 2008, Remarks on the House floor in support of H.R. 5811, the Electronic Message Preservation Act

(Cong. Rec. E1435)

A New Legal Term of Art Under the U.S. Federal Rules of Civil Procedure: *Electronically Stored Information* or "ESI"

"Electronically stored information":

- *The wide variety of computer systems currently in use, and the rapidity of technological change, counsel against a limiting or precise definition of ESI…A common example [is] email … The rule … [is intended] to encompass future developments in computer technology. --Advisory Committee Notes to Rule 34(a), 2006 Amendments*
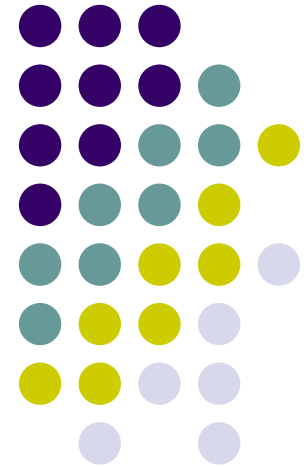
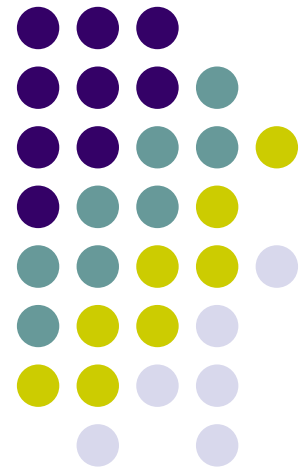# Email as ESI

- Email with or without Internet path metadata
- Email with internal audit trail metadata (who opened a message)
- Email with attachments (all kinds)
- Instant Messaging
- Text Messaging (from all devices)
- Unified Messaging Systems (voice mail on the desktop in .wav format)

Common Sources of ESI

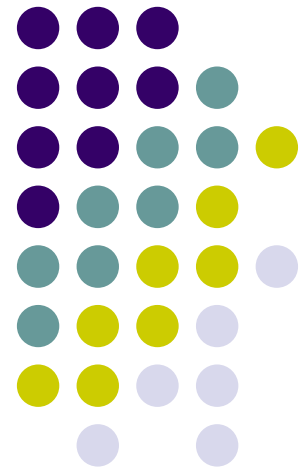Mainframes, network servers, local drives (including network activity logs)

DVDs, CD ROMs, floppy disks

Laptops

Backup tapes

External hard drives (e.g., flash, Zip, Jazz, ipods)

Third party storage
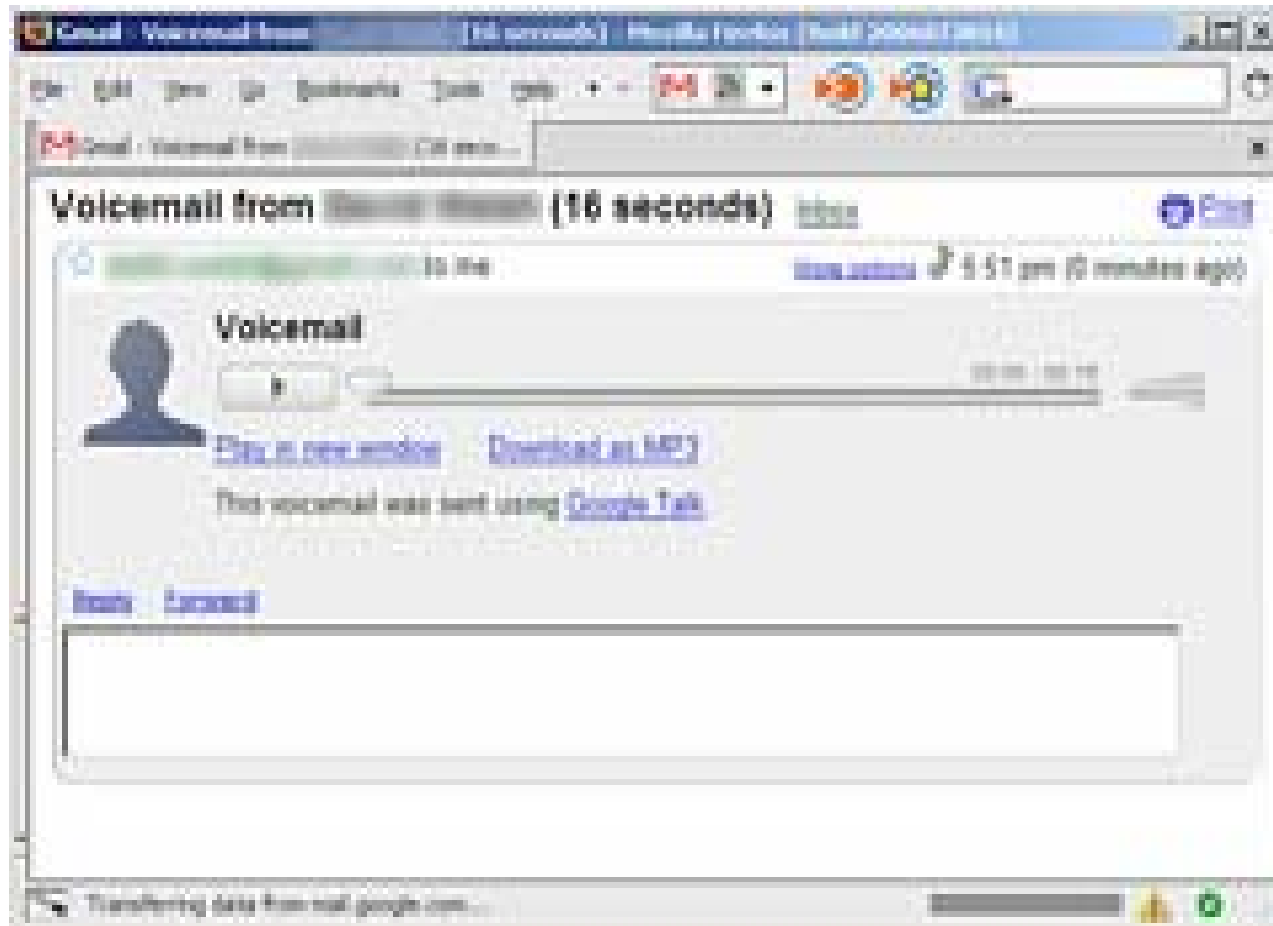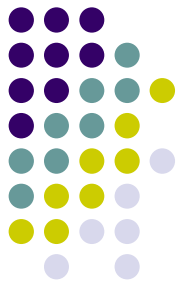
Hot topic: Metadata

- What is it?
  - Email header information (possibly hidden)
  - Proprietary features of word processing (e.g. summary fields)
  - Embedded & shadow data
  - Deleted keystrokes
  - Tracking info
  - Spreadsheet formulas
- Format issues and metadata
- Metadata ethics: inadvertent production
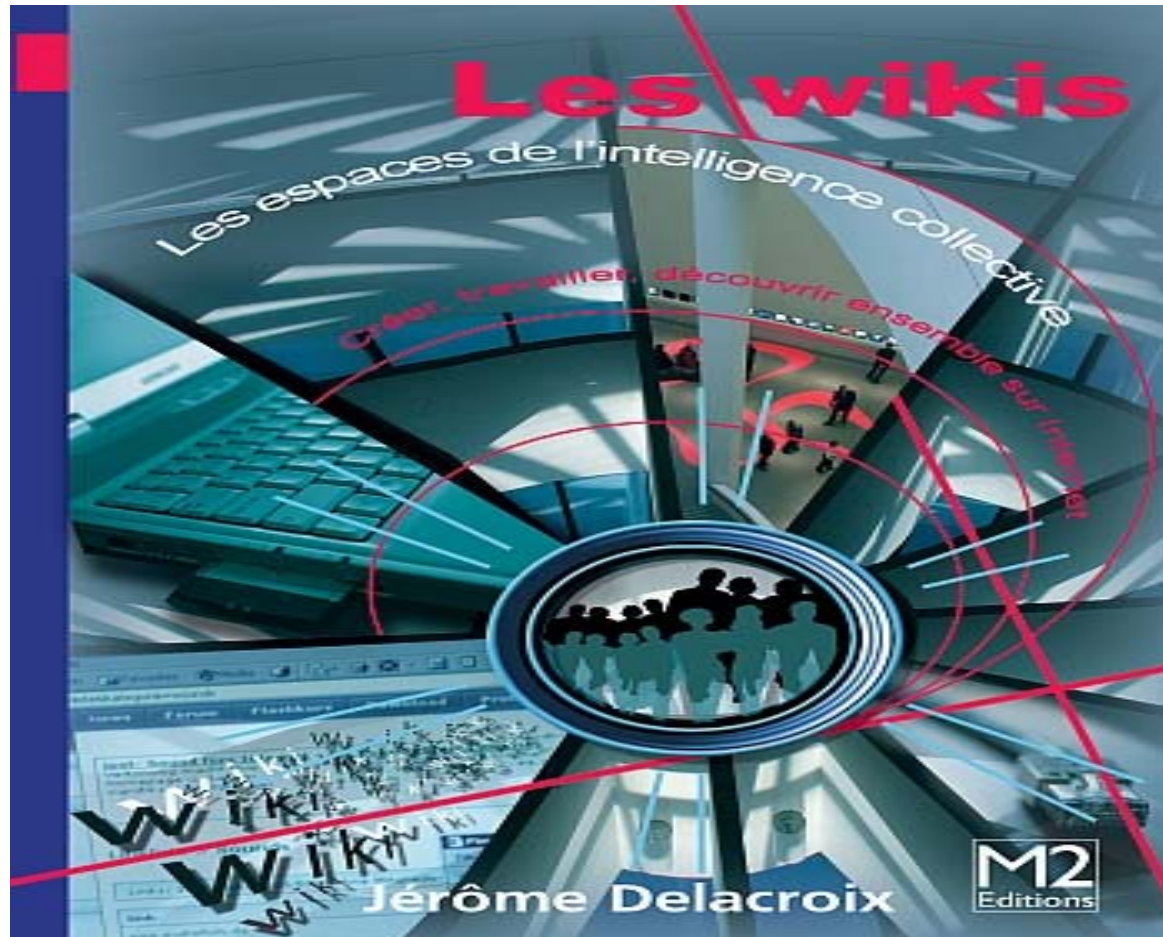
Text messaging, 2008-style

Voice mail

Social Software on the Web

Wikis
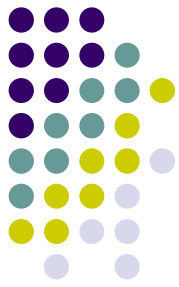
Blogs

Microblogs (e.g., twitter)

Virtual worlds

- The Library of Congress' virtual Declaration of Independence display as officially announced and which has opened as an Info Island in Second Life. The exhibit includes dioramas, streamed audio, text in the form of larger-than-life documents, information kiosks and even period furniture.

The Future: Public Records in the Clouds?

The New Reality: E-Discovery

Corporate Delete Keys Busy as E-Mail Becomes Star Witness at Trials

out of context and use it against you, and we have to guard against that."

Amazon, like many companies, had embraced E-mail as the preferred way to communicate about

posing new restrictions on E-mail.

This backlash underscores a paradox of the information age: While instantaneous electronic communication often helps productivity and innovation, the privacy of workplace correspondence and accountability for what is written are entering con
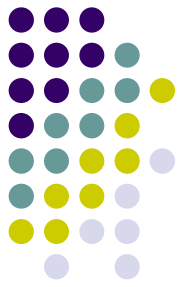
from their context.

But that is often not the case in legal disputes, where E-mail is increasingly being treated as the ultimate window into the true thoughts of executives and the inner workings of an enterprise.

signed for disk space management," said Jim Browning, a senior research analyst at the Gartner Group consulting firm. "The new question is how quickly should E-mail be deleted to prevent it from becoming a danger to the organization?"

As the 83 million Americans who now use E-mail in the workplace know, it is easily and often forwarded, copied and stored in perpetuity on everything from company backup tapes to personal Palm Pilots. The fact that people are disinclined to

that searches E-mail for offending words. If it finds one, a warning pops up before the E-mail is sent: "This is inappropriate language to use in correspondence with any customer, and a copy of this message will be immediately sent to your superior."

The Litigation Minefield

- **U.S. litigation increasingly demands the preservation of and access to all relevant documents, including in the form of "electronically stored information" or "ESI"**

- **Courts impose sanctions on parties for failing to preserve evidence under the "spoliation" doctrine**

- **Absent saving everything, often it is only with 20/20 hindsight that one can determine what *should* have been preserved in response to a lawsuit**

- **Recordkeeping solutions that rely on human judgment are prone to being second-guessed by litigants and judges.**

Two Cautionary Tales …

Qualcomm Inc. v. Broadcom Corp., 2007 WL 2296441(S.D. Cal. Aug. 6, 2007) (opinion recommending sanctions, involving underlying failure of a party to disclose 200,000 emails prior to trial)

United Medical Supply Co. v. United States 2007 WL 1952680 (Fed. Cl. June 27, 2007) (sanctions imposed for failure to adequately preserve ESI based on faulty email communications with contractors)
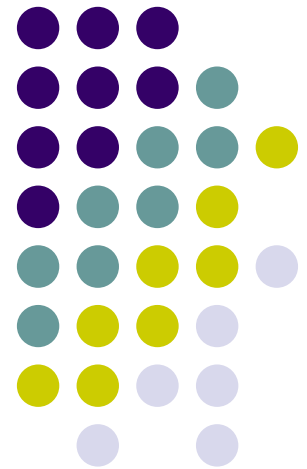
Armstrong v. EOP (filed 1989)

The Crisis: Last minute temporary restraining order (TRO) at end of Reagan Admin. & start of 7+ years of litigation

The Result: Injunctions, restoration of backups at huge expense, plus WH email archiving

Takeaways:

    -- recognized need to manage e-mail records

    -- WH email archiving possible with record tagging

    -- importance of metadata

    -- legacy issue of backup tapes

*Armstrong v EOP*, 1 F.3d 1274 (D.C. Cir. 1993)
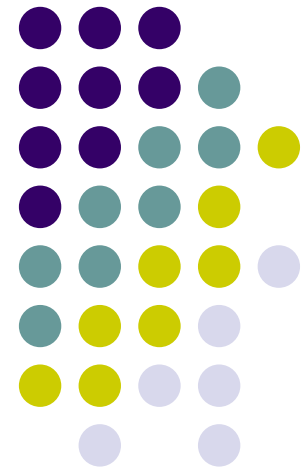
Public Citizen v Carlin (1998)

The Crisis: Threat of "no delete" rule applied government-wide while thousands of records schedules reviewed

The Result: General Record Schedule 20 upheld on appeal

Takeaway: GRS 20 & 24 now allow email and other forms of ESI to be deleted and/or recycled assuming recordkeeping obligations are otherwise met
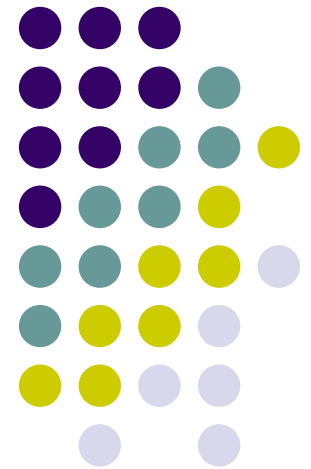
Public Citizen v. Carlin, 184 F.3d 900 (D.C. Cir. 1999), *cert. denied,* 529 U.S. 1003 (2000)
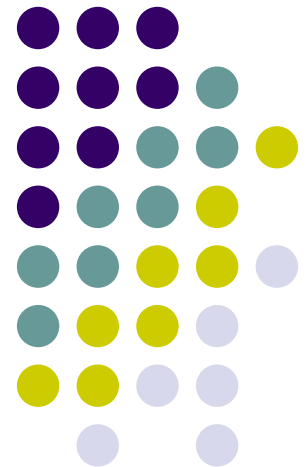
Alexander v. FBI (2000)

The Crisis: Missing WH email due to technical issues with email archiving

The Result: Huge restoration project for EOP backup tapes + GAO investigation
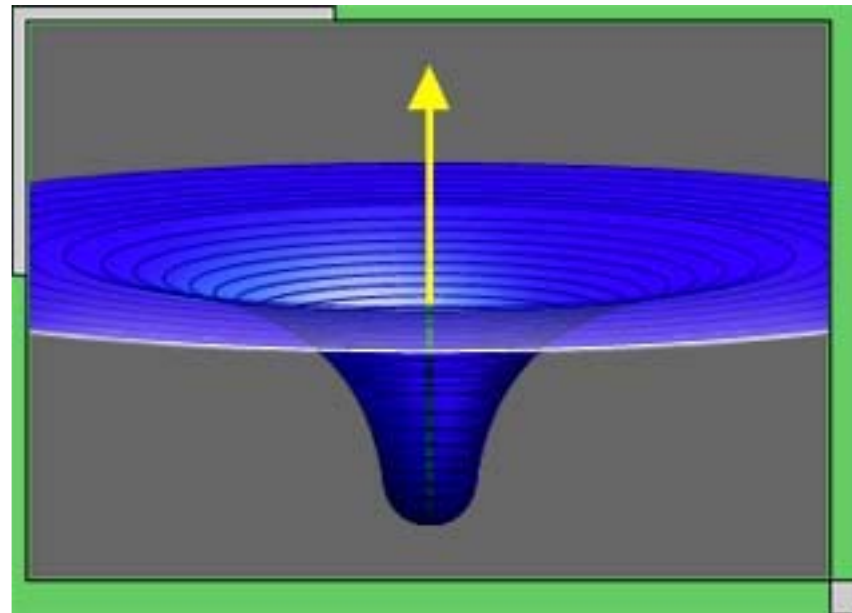
Takeaway: QC measures for electronic records a must

*Alexander, et al. v. FBI, et al.*, 541 F. Supp. 2d 274 (D.D.C. 2008)

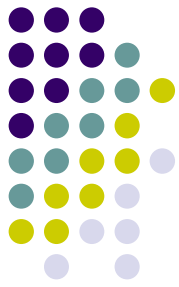# Information Inflation: The Expanding ESI Universe . . . .

Foundational elements of recordkeeping in the federal government

+ Agency file plans and agency records schedules

+ Appraisal decisions signed off by Archivist on what constitute permanent records (eventually accessioned into NARA), and what constitute temporary records

(stored by agencies during active use and then at offsite federal or other record centers for the duration of the retention period of the records)

+ Record schedules subject to public notice in Federal Register

+ General Records Schedules for admin. records

See 44 USC 3303, 3303a(a), (d)
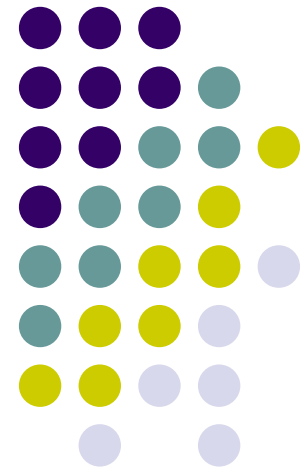
Lifespan of "Temporary" Records

- The definition of what constitutes a federal record (44 USC 3301) allows for tremendous flexibility in what is considered to be "record" material, spanning from ephemeral records (retained for hours/days) thru to short term temporary (weeks/months), long term temporary (years or decades) to permanent records (forever)

- Problem: matching up the retention span of various e-records with the working life of information technology

# Federal agency recordkeeping as applied to email

- All federal records are retained as part of "file plans" and "record retention schedules," catalogued as discrete "record series" in accordance with retention and disposition instructions

- High degree of granularity present in records schedules as to specific retention periods for temporary records

- Appraisal by archivists is primarily a matter of justifying the segregation of "wheat" and "chaff," i.e., the permanent from the temporary

- Email & ESI appraised under traditional methods: expected to be segregated by record series.
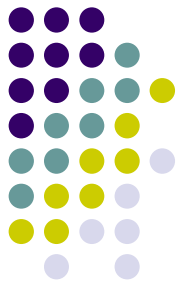
Transitory Email Reg

- NARA final regulations published in the Federal Register on February 21, 2006 (71 F.R. 8806), modifying 36 CFR 1234.24

- Email records appropriate for preservation for less than 180 days may be managed on live email systems and allowed to be deleted as part of automatic processes, without a user further needing to print out or electronically archive.
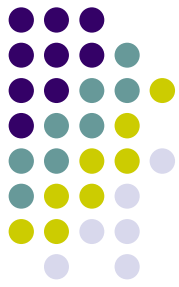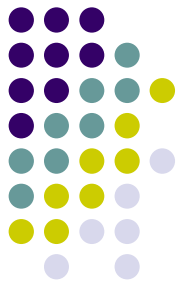
H.R. 5811: EMPA

- Electronic Message Preservation Act, sponsored by Rep. Henry Waxman, passed the House on July 9, 2008
- Would have Archivist issue regulations
  - Requiring electronic capture, management and preservation of electronic records
  - Requiring electronic records to be readily accessible for retrieval through electronic searches
  - Establishing mandatory minimum functional requirements for electronic records management systems to ensure compliance

H.R. 5811: EMPA (cont'd)

- EMPA would require NARA to promulgate regulations within 18 months of enactment
- EMPA would require regulations to include timelines for agencies "that ensure compliance as expeditiously as practicable but not later than four years after the date of enactment"
- EMPA would require that, to the extent practicable, the regulations include requirements for the capture, management, and preservation of other forms of electronic records beyond "electronic messages"
- Reporting requirements for agencies: 4 years after enactment

H.R. 5811: EMPA (cont'd)

- "Electronic messages" means electronic mail and other electronic messaging systems that are used for purposes of communicating between individuals
- "Electronic records management system" means a software system designed to manage electronic records with an information technology system, including by
  - (A) categorizing and locating records;
  - (B) ensuring that records are retained as long as necessary
  - (C) identifying records that are due for disposition;
  - (D) the storage, retrieval and disposition of records.
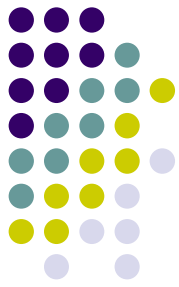
Four Prior Recordkeeping "Paths"

- **Print to paper (GRS 20, Item 14)**
- **Backup tapes --** *but see* **36 CFR 1234.24(c) (backup tapes should not be used for recordkeeping purposes)**
- **Online user-based foldering in proprietary live systems**
- **Electronic recordkeeping under Records Management Applications (RMAs)**

Impact of Technology on E-Records Management Applications: Snapshot 2008

- **A universe of proprietary products exists in the marketplace: document management and RMAs**
- **DoD 5015.2 compliant products**
- **However, scalability issues exist**
- **Utopia is records mgmt without extra keystrokes**
- **Agencies must prepare to confront significant front-end process issues when transitioning to electronic recordkeeping**
- **Records schedule simplification is *key***

Electronic Archiving

- What is it?

  100% snapshot of (typically) email, plus in some cases other selected ESI applications

- How does it differ from an RMA?

  Goal is of preservation of evidence, not records management *per se*

- NARA Bulletin 2008-05

# The Future "Promise" of Total E-record Archiving

- **100% archiving of email & ESI on the desktop**
- **Transport out of email store into generic format (e.g., XML)**
- **Use of smart filter technologies to segregate permanent from the temporary**
- **Culling for non-record material using certain agreed-upon protocols**
- **Default temporary record status of remaining archived materials**
- **However: all "eggs" in the search basket**

The Myth of Search & Retrieval

When lawyers (and others) request production of "all" relevant documents (and now ESI), all or substantially all will in fact be retrieved by existing manual or automated methods of search.

Corollary: in conducting automated searches, the use of "keywords" alone will reliably produce all or substantially all documents from a large document collection.
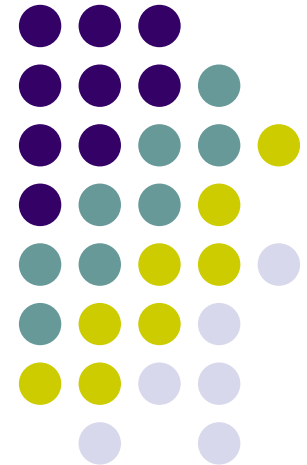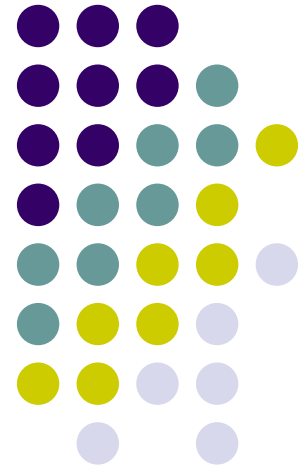
The "Hype" on Search & Retrieval

Claims in the tech sector that a very high rate of "recall" *(i.e., finding all relevant documents) is easily obtainable provided one uses a particular software product or service.

Corollary: claims that documents can be easily segregated by examination of content.

The Reality of Search & Retrieval

+ Past research (Blair & Maron, 1985) has shown a gap or disconnect between lawyers' perceptions of their ability to ferret out relevant documents, and their actual ability to do so:

--in a 40,000 document case (350,000 pages), lawyers estimated that a manual search would find 75% of relevant documents, when in fact the research showed only 20% or so had been found.
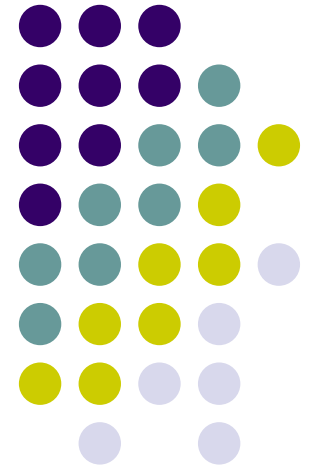
More Reality: IR is Hard

+ Information retrieval (IR) is a hard problem: difficult even with English-language text, and even harder with non-textual forms of ESI (audio, video, etc.) caught up in litigation.
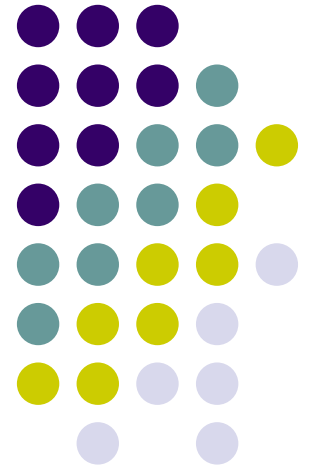
+ A vast field of IR research exists, including some fundamental concepts and terminology, that everyone -- not just lawyers -- would benefit from having greater exposure with.
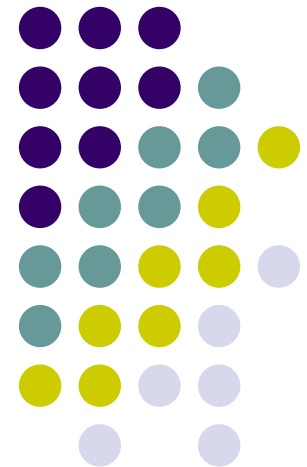
Why is IR hard (in general)?

+ Fundamental ambiguity of language
+ Human errors
+ OCR problems
+ Non-English language texts
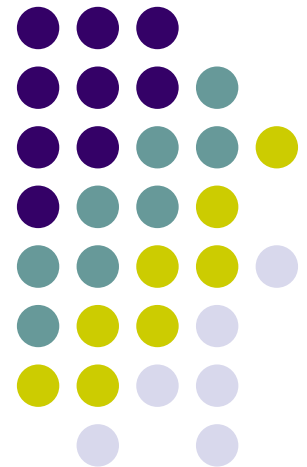**+ Nontextual ESI (in .wav, .mpg, .jpg formats, etc.)**
+ Lack of helpful metadata

Problems of language

Polysemy: ambiguous terms (e.g., "George Bush," "strike,")

Synonymy: variation in describing same person or thing in multiplicity of ways (e.g., "diplomat," "consul," "official," ambassador," etc.)

Pace of change: text messaging, computer gaming as latest examples (e.g., "POS," "1337")

Judge Grimm writing for the U.S. District Court for the District of Maryland

"[W]hile it is universally acknowledged that keyword searches are useful tools for search and retrieval of ESI, all keyword searches are not created equal; and there is a growing body of literature that highlights the risks associated with conducting an unreliable or inadequate keyword search or relying on such searches for privilege review." *Victor Stanley, Inc. v. Creative Pipe, Inc.,* --- F.Supp.2d ----, 2008 WL 2221841, * 3 (D. Md. May 29, 2008); *see id., text accompanying nn. 9 & 10* (citing to Sedona Search Commentary & TREC Legal Track research project)

Judge Facciola writing for the U.S. District Court for the District of Columbia

"Whether search terms or 'keywords' will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics. *See* George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?', 13 RICH. J.L. & TECH.. 10 (2007)* * * * Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread."
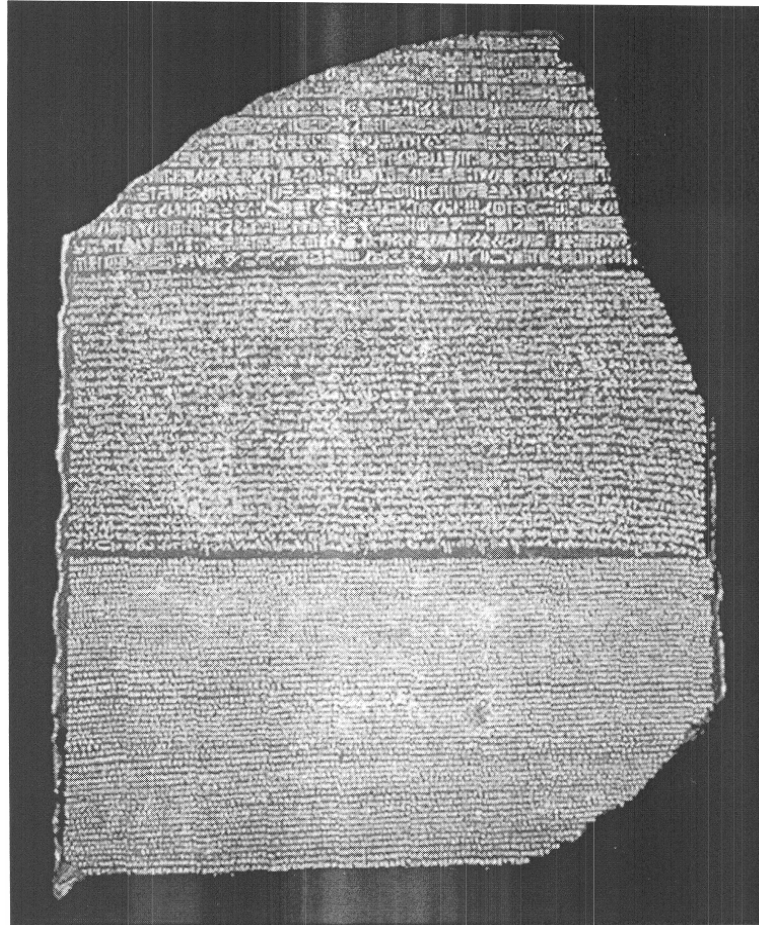
-- ***U.S. v. O'Keefe***, 537 F.Supp.2d 14, 24 D.D.C. 2008).

Beyond Boolean: getting at the "dark matter"

(i.e., relevant documents not found by keyword searches alone)
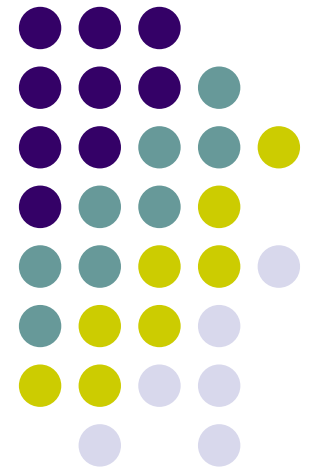
Rosetta Stone Approach: The Need To Master 3 Languages: Legal, RM, IT
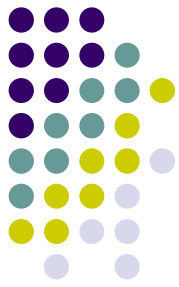
In The Age of Email Archiving…

- The IT savvy lawyer, records manager, and archivist will all be able to crosswalk between technological solutions for filtering exponentially increasing volumes of information.

- The "information retrieval" savvy lawyer and access professional will understand that a range of alternative search methods – including AI methods yet to be developed -- hold the *key* to the efficient future accessing of information

- Lawyers and records managers and IT staff will strive to overcome the difficulty of failing to share a common language.

Best practices in managing ephemeral records in the form of electronically stored information: back to the fundamentals

* Inventorying for the purpose of obtaining intellectual control over your Agency's electronic systems to account for forms of ephemeral records (network applications, backups & legacy media)

* Understanding existing records schedules & retention periods for your Agency's records (and consider updating & simplifying schedules in anticipation of transition to electronic recordkeeping)

* Consider appointing "Knowledge Counsel" in General Counsel and Solicitor offices to act as agents of change, working with CIOs and records officers

* Formulating explicit e-records guidance on what constitute records, and including holds guidance

* Providing training on FRA and e-discovery obligations

Additional Resources

- **Title 44 U.S. Code, Chaps 21, 29, 31, 33**

- **36 C.F.R. Part 1234 E-Records Mgmt**

- **http://www.archives.gov/records-mgmt/ (comprehensive records management website, with bulletins and a toolkit of best practice white papers, including on transitioning to electronic recordkeeping; FAQs on many subjects, including scanning documents, instant messaging, wikis, RSS feeds, and email archiving, for the latter see NARA Bulletin 2008-05)**

- **www.thesedonaconference.org (Sedona white papers, including the Sedona Principles 2d ed. 2007, plus papers on Email management, on Search and Information Retrieval in E-discovery, and many more)**
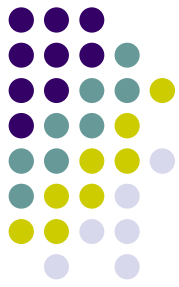
Jason R. Baron

Director of Litigation

Office of General Counsel

National Archives and Records Administration

(301) 837-1499
Email: jason.baron@nara.gov