



# Food For Thought

Applying Civil Litigation  
eDiscovery Best Practices

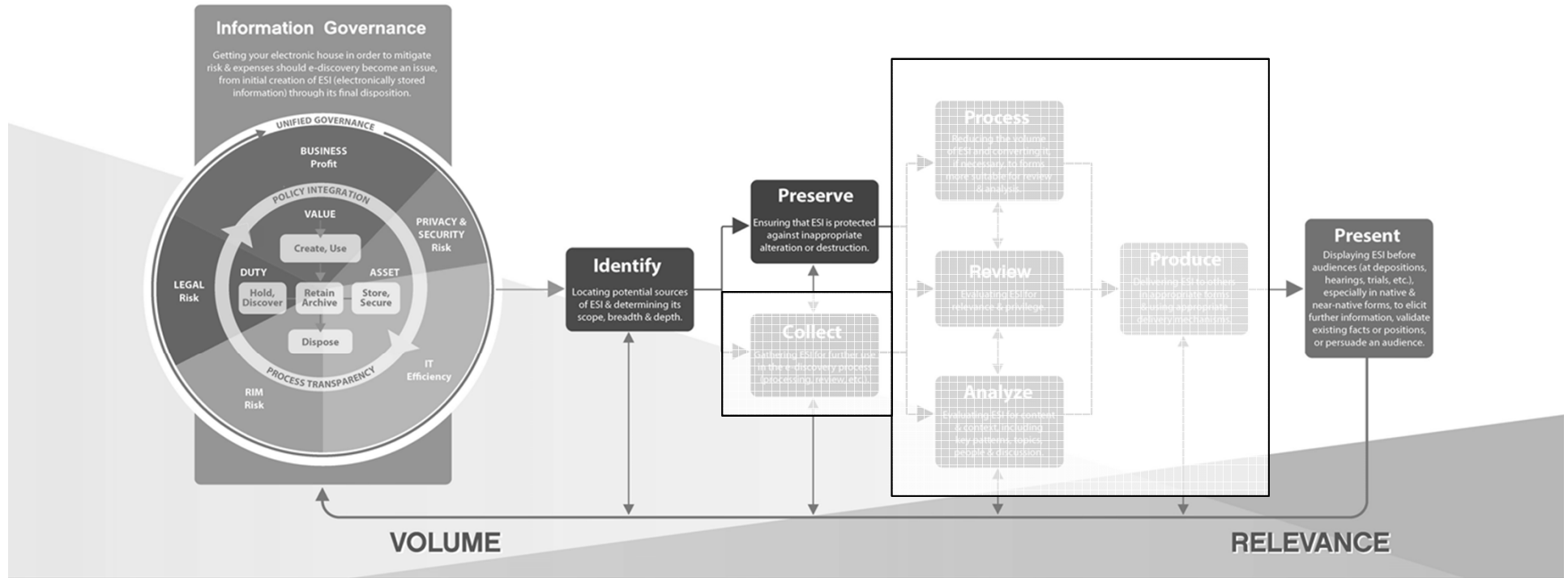
# Why Does This Matter?

- Hillary Clinton Emails
- Mueller/Manafort Redactions
- March 2018: US sets new record for censoring, withholding gov't files\*
  - FOIA responses: censored files or no response in 78% of 823,222 requests
  - \$40.6mm defending decisions

\*<https://apnews.com/714791d91d7944e49a284a51fab65b85>

# Electronic Discovery Reference Model

Standards, Guidelines, and Practical Resources for Legal Professionals and E-Discovery Practitioners



Founding Sponsors



Gold Sponsors



Silver Sponsors



**E·D·R·M**  
**DUKE LAW**

EDRM at Duke Law School  
210 Science Drive Box 90362  
Durham, NC 27708-0362  
EDRM@law.duke.edu

[www.edrm.net](http://www.edrm.net)

Oct 2018



# Collections

- Scoping is essential
  - Custodians
  - Sources
  - Locations
  - Timeframes
- Some available technologies
  - Office 365
  - Document Management Systems
- Over-collection
  - Larger volumes to search, review, and produce
  - Increased costs (\$\$\$)

# Search - Traditional Principles

- Under Rule 34, courts do not control how search is done
- Usual rule is that courts are agnostic as how search is done
- Intervene only when there is showing of deficiency
- “There must be more” or “we found an email they missed” are insufficient

## Sedona Principle #6

- Responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information.

# Approaches

1. Exhaustive Manual Review
2. Keyword Searches
3. Technology Assisted Review

# Manual Review

- 1 billion emails, 25% have attachments
  - 50 documents per hour
  - 100 reviewers, 10 hours per day, 7 days per week, 52 weeks per year
- 54 years / \$2 billion
- 100,000 documents
  - \$3 to \$4 per document to review (1<sup>st</sup> level, 2<sup>nd</sup> level, priv, etc.)



# Manual Review – Gold Standard?

- TREC 2009 Legal Track
  - Manual reviewers missed 21% to 75%
- Deficiency examples
  - Disagreement among reviewers as to whether the document is relevant
- Can it be measured?

# Keyword Searching

- Phrases and words most likely to be found in responsive documents
- Process
  - Create index of the document text
  - Apply keywords, review search term report
  - Review documents
- Sophistication (Boolean, Proximity, etc.)

# Keyword Searching

- Search Term Report tells us nothing about Responsiveness
  - Term A: 10,000 documents
    - 99% non-responsive
    - 9,999 documents that DON'T MATTER reviewed
- Abbreviations / Acronyms
- “Bill”
  - Invoice?
  - Legislation?
  - First Name?

## United States v. O'Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008)

- Judge orders government to produce requests for visa applications submitted by the defendant.
- Government used “early or expedite” or “appointment or early & interview” or “expedite & interview.”

# United States v. O'Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008)

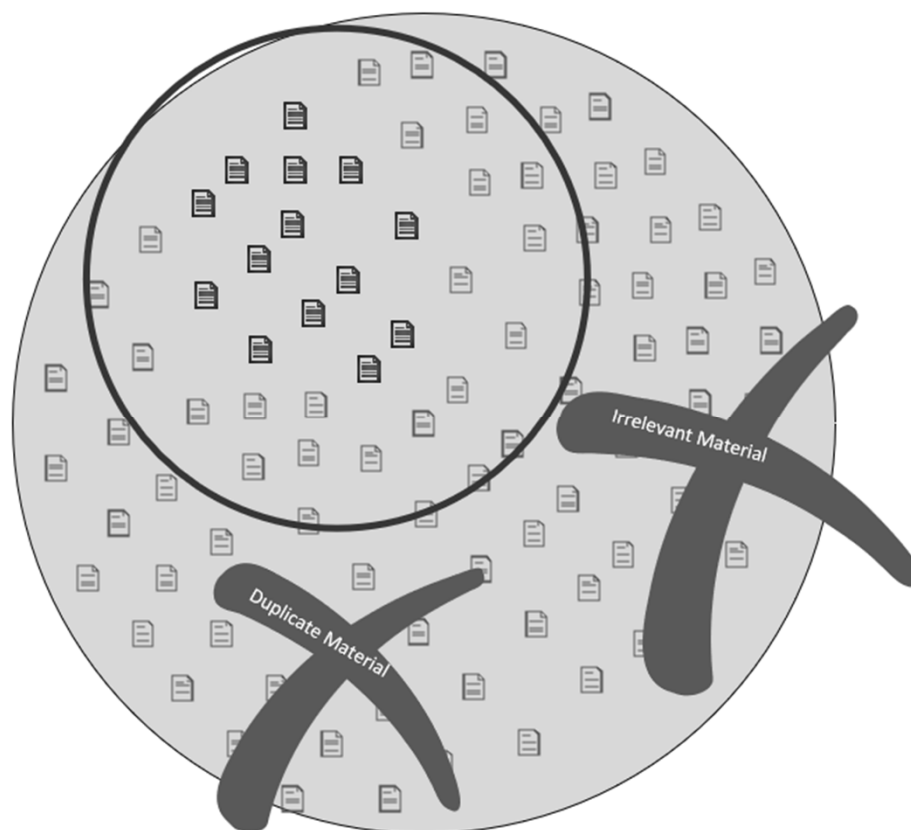
- Whether search terms or "keywords" will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics.... Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread.
- This topic is clearly beyond the ken of a layman and requires that any such conclusion be based on evidence that, for example, meets the criteria of Rule 702 of the Federal Rules of Evidence. Accordingly, if defendants are going to contend that the search terms used by the government were insufficient, they will have to specifically so contend in a motion to compel and their contention must be based on evidence that meets the requirements of Rule 702 of the Federal Rules of Evidence.

# Technology Assisted Review (“TAR”)

- Review workflow that incorporates technologies, including:
  - Email threading
  - Near duplicate detection
  - Machine Learning
- Also known as: TAR, CAL, Predictive Coding, etc.

# TAR – Reduce the Burden of Review

- Find documents not found with basic keywords
- Items grouped logically to speed review
- Identify Relevant documents quickly
- Minimize review of Non-Relevant documents



# TAR - Considerations

- What do you want to ensure the other side provides you?
  - What are you willing to provide?
- Are you agreeing to things outside of your control?
- What are you validating and what is an acceptable level?
- Are you meeting the obligations you agreed to?
- Are you using TAR in a way that does not need to be disclosed?



Rio tinto v. vale, S.A., 2015 WL  
4367250 (S.D.N.Y. July 15, 2015)

- While court prefers disclosure, it concluded that sharing training sets is not necessary to ensure appropriate training of the model
- CAL renders seed set much less significant because the algorithm is constantly retrained as the algorithm identifies documents as potentially relevant or not.

Progressive Cas. Inc. Co. v. Delaney, 2014 WL 3563467  
(D.Nev. July 18, 2014)

- Search terms used under agreed protocol yielded 565,000 hits of 1.6 million documents. Plaintiffs propose to use TAR on the “hits” claiming recall of 70% to 80% and then manually review what TAR yields.
- Denied because (1) plaintiffs refused to disclose seed sets; (2) unilateral switch after months spent by defendants narrowing search terms at plaintiffs’ insistence; (3) too late and plaintiffs has missed deadlines for production already
- Court noted result would have been different had parties agreed to protocol using TAR when they began

# TAR - Approval

- Da Silva Moore, express approval of TAR protocol when 3 million documents needed to be reviewed
- Studies have shown superior to manual review
- Encourages transparency and cooperation; if cannot agree, court approval is an option
- Perfection is never required; only reasonableness

# TAR – Can it be Ordered?

- Hyles v. New York, 2016 WL 407714 (S.D.N.Y. Aug. 1, 2016)
  - Sedona principle 6 cited
  - Standard is reasonableness, not perfection
- Kleen Products, LLC v. Packaging Corp., 2012 WL 4498465 (N.D. Ill. Sept. 28, 2012)
  - Plaintiff sought to require defendants to abandon Boolean search (which had cost \$1million) and switch to TAR. Motion denied, citing Sedona Principle 6 but parties ordered to meet and confer.
  - Ultimately agree that defendant not re-do while reserving right to attack sufficiency and propose alternative search methods for future productions.
- In re: Bridgeport Edu. Inc. Sec. Litig, 2014 WL 3867495 (S.D. Cal. Aug. 6, 2014)
  - Motion that defendants be compelled to use TAR denied when defendants had already searched the collection with search terms

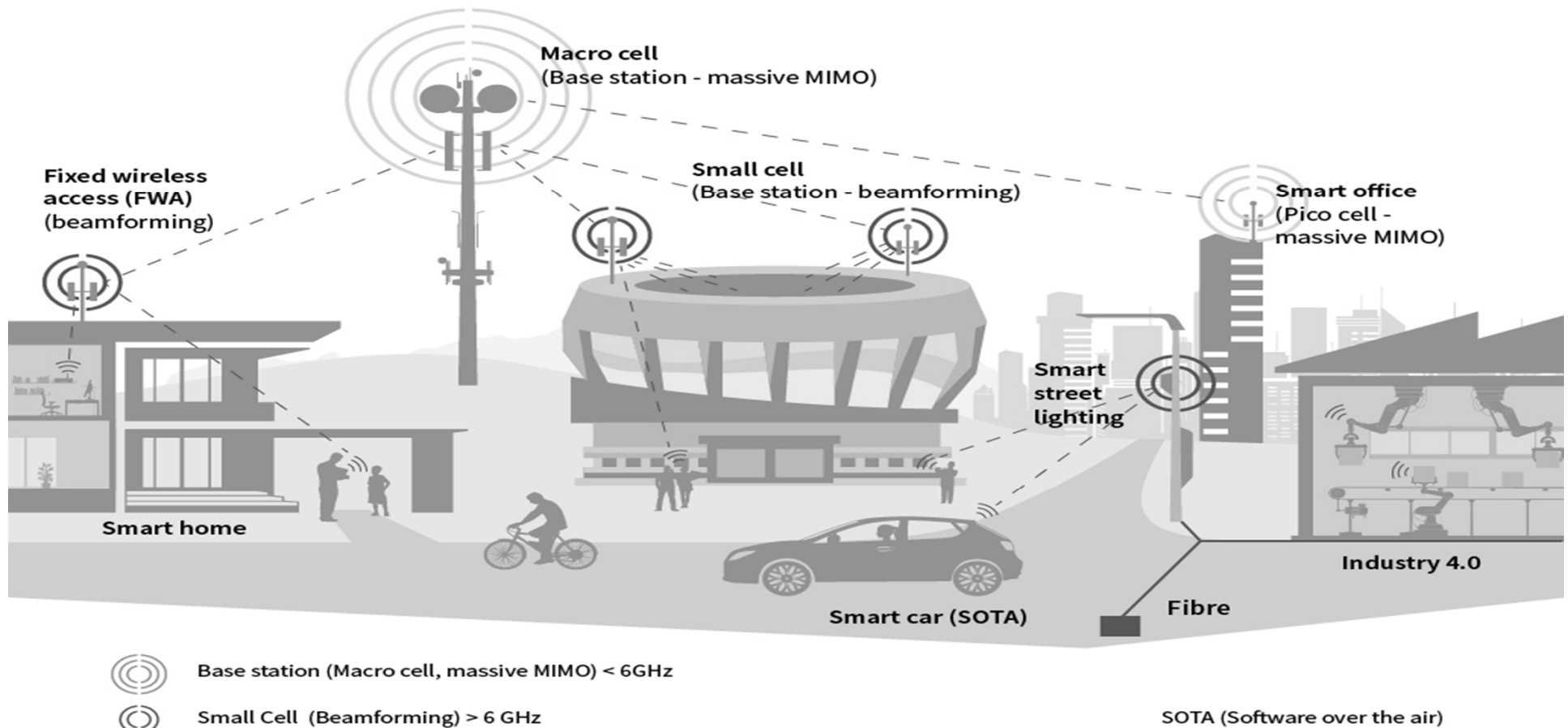
# TAR + Keywords

- In re Biomet M2A Magnum Hip Implant Products Liability Litigation, 2013 WL 1729682 and 2013 WL 6405156 (N. D. Ind. April 18 & 21).
  - Keywords used to cull documents from 19.5 to 3.9 and then used TAR on subset.
  - Plaintiff complains that use of keywords before TAR tainted the result and defendants should use TAR on 19.5 documents.
- REJECTED: “It might well be that predictive coding instead of a keyword search would unearth additional relevant documents. But it would cost Biomet a million, or millions, of dollars to test the plaintiff’s theory that predictive coding would produce a significantly greater number of relevant documents. Even in light of the needs of the hundreds of plaintiffs in this case, the very large amount in controversy, the parties resources, the importance of the issues at stake, and the importance of this discovery in resolving the issues, I can’t find that the likely benefits of the discovery proposed by plaintiff equals or outweighs its additional burden and an additional expense to Biomet.”

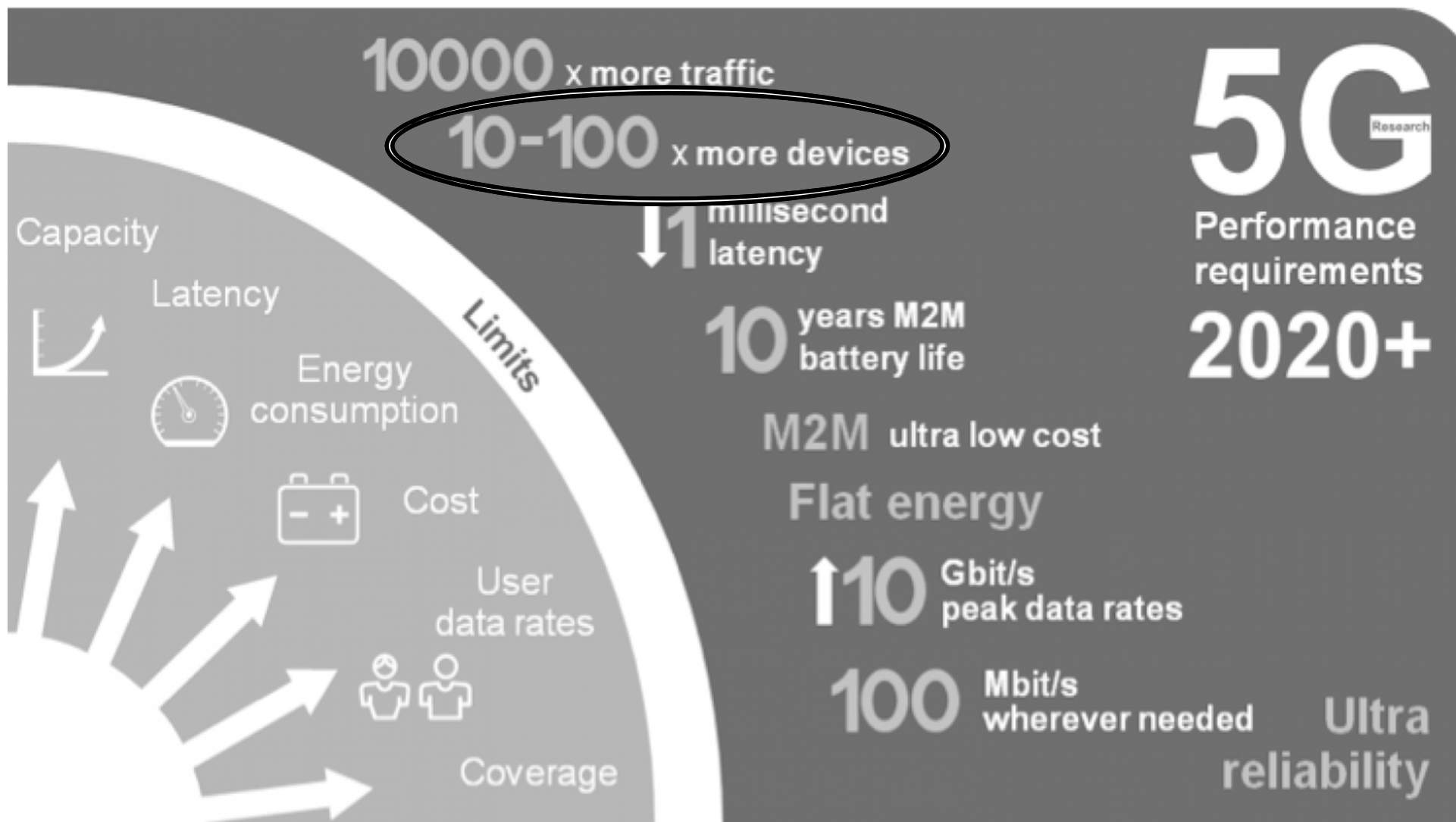


What is  
Big Data?

## Smart and connected - the communication of tomorrow with 5G



iDiscovery Solutions







# Insight to Connections and Patterns

Don't worry. This will NEVER  
affect the practice of law



# Fitness Trackers

Richard Dabate told police a masked intruder assaulted him and killed his wife in their Connecticut home. His wife's Fitbit told another story and Dabate was charged with the murder.

**POLICE LINE DO NOT CROSS**

# Smart Meters

James Bates said an acquaintance accidentally drowned in his hot tub in Arkansas. Detectives suspected foul play and obtained data from Bates's Smart Water Meter. Bates was charged with murder.

**POLICE LINE DO NOT CROSS**

# Pacemaker

Ross Compton told investigators he woke up to find his Ohio home on fire and climbed through a window to escape the flames. Compton's pacemaker suggested otherwise. He was charged with arson and insurance fraud.

The image features a complex network graph visualization. It consists of numerous circular nodes of varying sizes, interconnected by a dense web of thin, light-colored lines representing edges. The nodes are distributed across the frame, with some appearing as central hubs connected to many other nodes, while others are more isolated. The background is a dark, textured gray. A semi-transparent horizontal band spans the lower portion of the image, containing the text. The right side of this band features a fine grid pattern.

The Data Knows Us...  
Better Than We Know Ourselves



3000 K St. NW, Suite 330  
Washington, DC 20007



202.249.7860



[idiscoverysolutions.com](http://idiscoverysolutions.com)



@iDiscoveryInc

# Thank you!

Julian Ackert – [jackert@idsinc.com](mailto:jackert@idsinc.com)