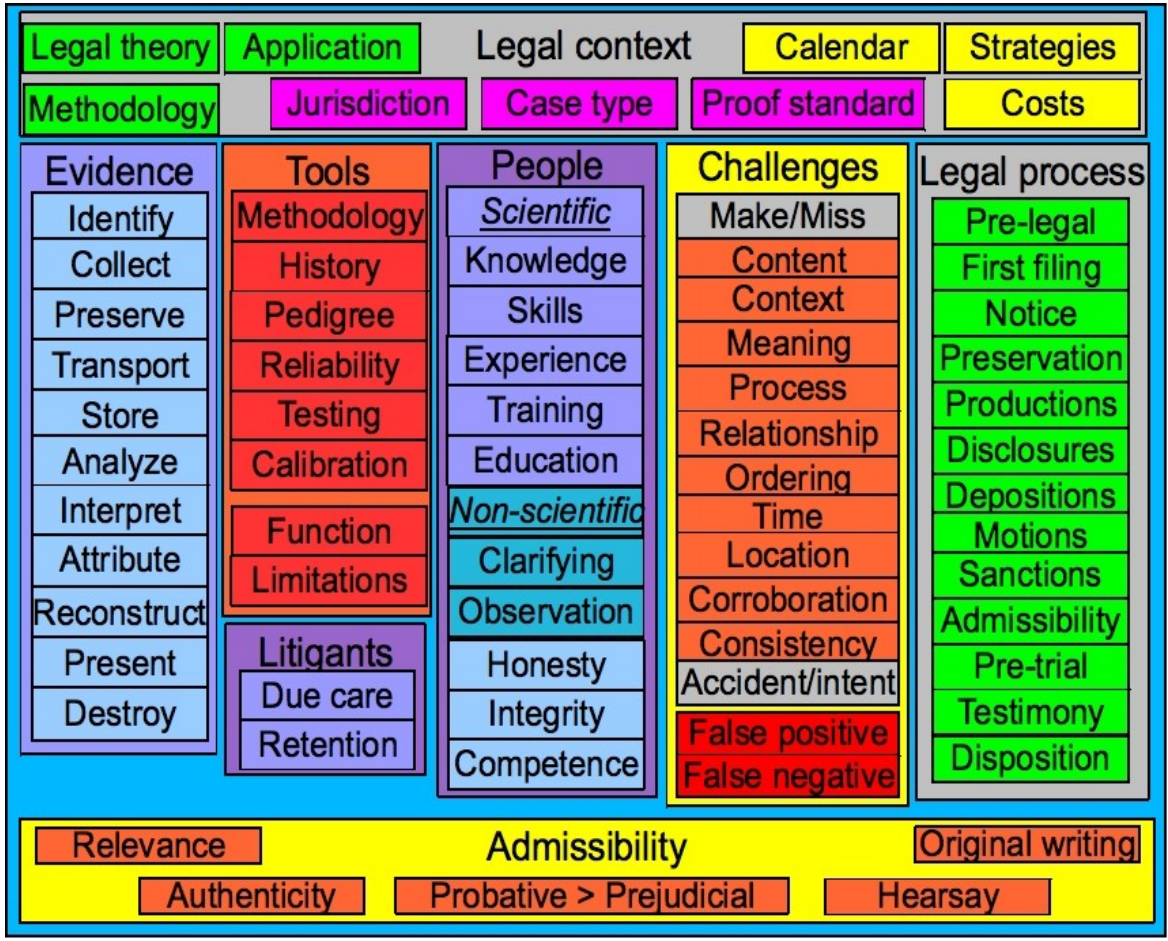
Recent and Hoped for Advances in Digital Forensics

Naval Postgraduate School – Aug 19, 2010

Dr. Fred Cohen
President - California Sciences Institute
CEO – Fred Cohen & Associates

Outline

- Background of the speaker and subject
- Building a science
- My current approach
- Limitations
- The long view
- Your turn!



Your speaker

- Education:
  - B.S. Electrical Engineering (C-MU '77)
  - M.S. Information Science (Pitt '81)
  - Ph.D. Electrical Engineering (USC '86)
- Experience:
  - >30 years of information protection R&D, design, engineering, testing, implementation, and operation
  - >20 years since first digital forensics case
- CEO - Fred Cohen & Associates
  - Enterprise information protection architecture
  - Digital forensics for high-valued legal cases

CalSci

- President – California Sciences Institute
  - Started doctoral classes in 2010-07
- M.S. And Ph.D. Program in National Security
  - Technical aspects of these fields
- M.S. In Advanced Investigation
- Ph.D. In Digital Forensics
  - The first Ph.D. program in Digital Forensics in the United States
- calsci.org

The challenge

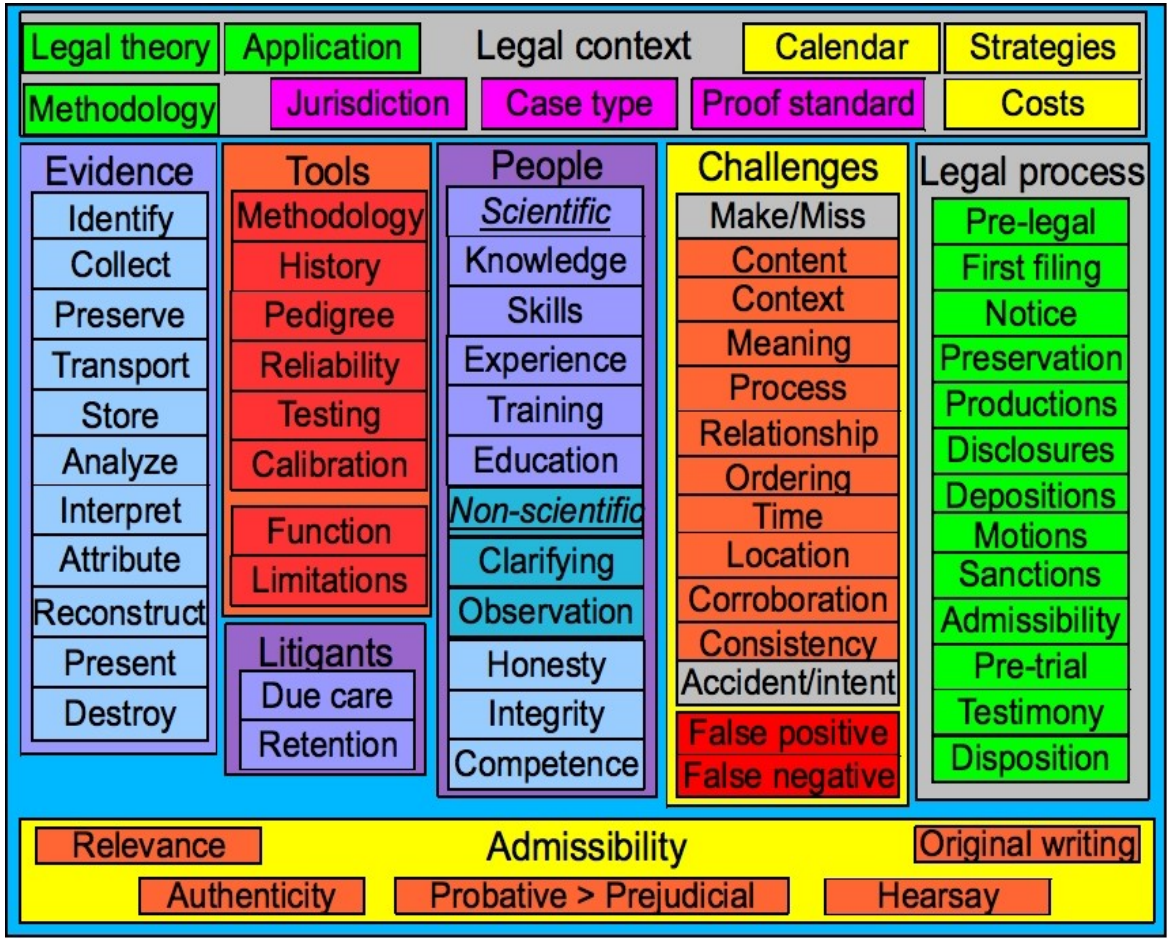
- Digital Forensic Evidence (DFE) Examination
  - Is not operating as “normal science”
    - There is only very limited community consensus
    - Terms are not well defined and consistently used
    - We don't have a well understood epistemology
    - We don't have widely used theory / methodology
    - We don't have a strong experimental basis
    - We don't have an agreed-upon physics
  - This we must change
    - By creating a community consensus
    - By defining and using terms consistently
    - By agreeing on an epistemology, theory, methodology, experimental basis, and physics

Law requires scientific basis

- Legal mandates (FRE) indicate that:
  - For expert testimony to be admitted
  - The expert must reliably apply a method accepted in the technical community
- However:
  - National Research Council (2009)
  - DoJ report (2005)
  - And others…
  - To sum up: forensics often lacks the scientific basis for its claims and many cases are overturned because of this

Outline

- Background of the speaker and subject
- Building a science
- My current approach
- Limitations
- The long view
- Your turn!



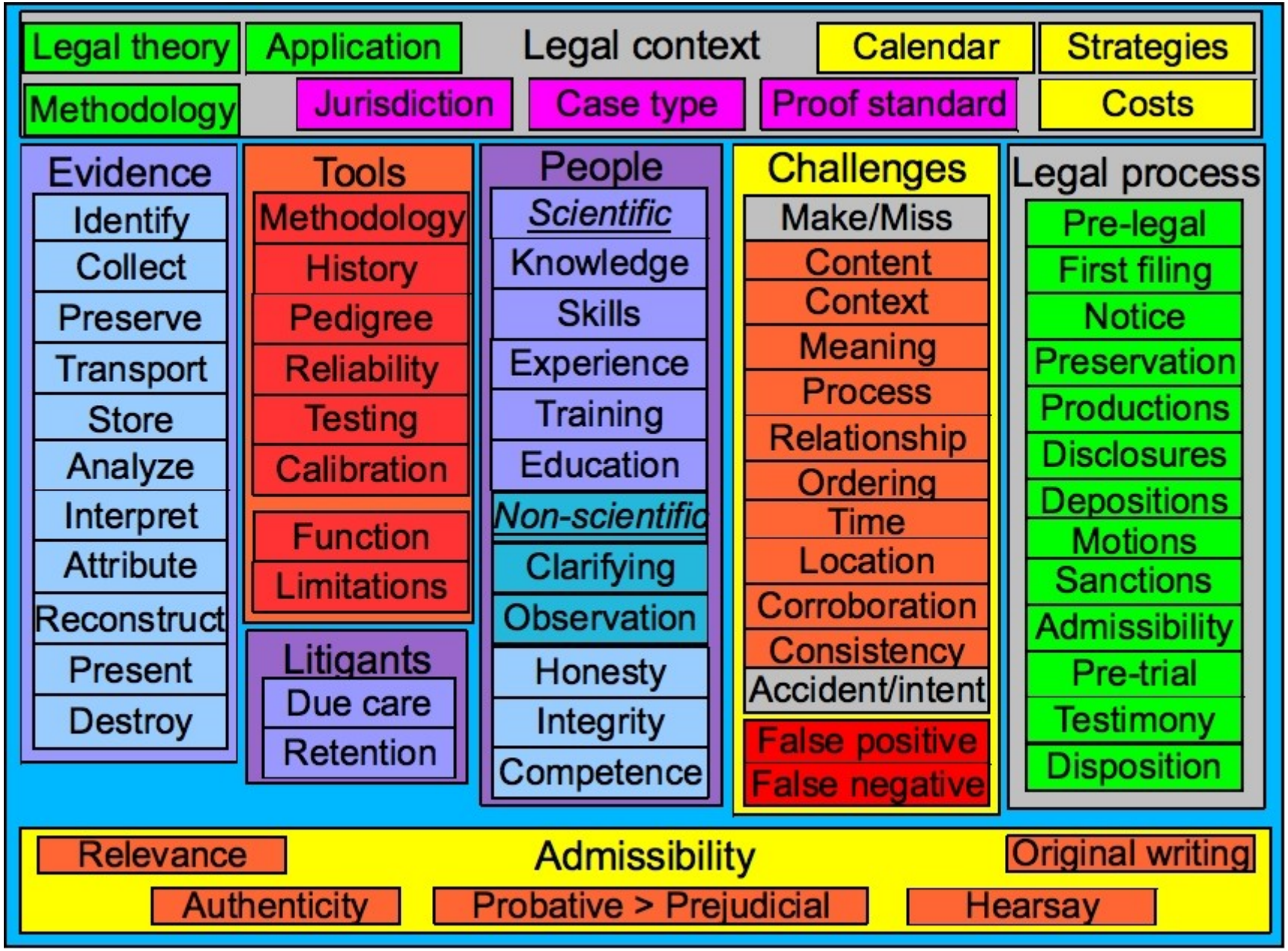
A dry subject?

- But the desert also has great beauty

Raise your hand if...

- Did any of you ever create a science?
  - If so... HELP!!!
- What is involved in doing such a thing?
  - The scientific method... whatever that is...
  - And then what?
- My view is to start with history
  - How did others do this?
  - How are things different today?
- The big picture today

The big picture today



The issues at hand

- Legal theory
- Methodology
- Application

- Knowledge
- Skills
- Experience
- Training
- Education

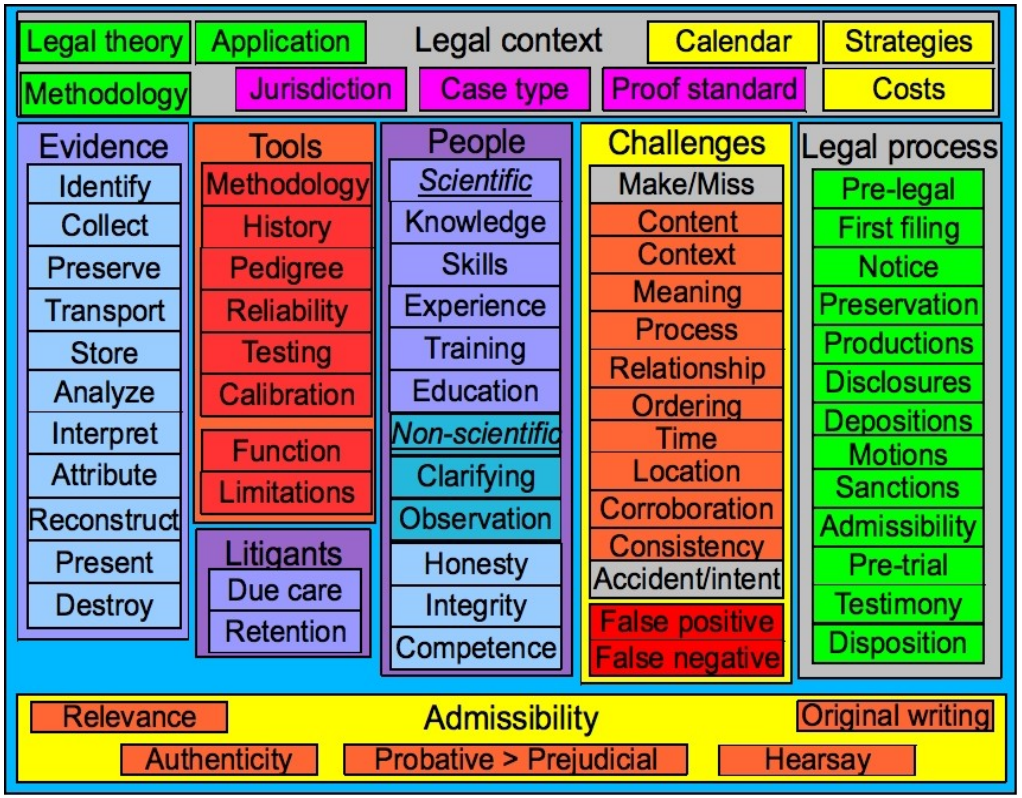
Examination:
- Analyze
- Interpret
- Attribute
- Reconstruct
- (Present)

- Methodology
- History
- Pedigree
- Reliability
- Testing
- Calibration

- Functions
- Limitations

Outline

- Background of the speaker and subject
- Building a science
- My current approach
  - Epistemology
  - Theory
  - Methodology
  - Experimental basis
  - Information Physics
- Limitations
- The long view
- Your turn



Epistemology

- The branch of philosophy that studies the nature of knowledge, its presuppositions and foundations, and its extent and validity.
- In the case of the science of digital forensic evidence examination, this implies:
  - Digital evidence is entirely sequences of bits.
  - Physics different than matter and energy.
  - Finite granularity.
  - Observation without alteration.
  - Duplication without removal.
  - Finite, but short, times.

Epistemology 2

- All DFE is trace, but not transfer.
- DFE is normally latent in nature.
  - It can only be observed through the use of tools.
  - → many issues with respect to those tools.
- DFE is produced by FSMs.
  - FSMs have specific properties that define a portion of the physics of DFE.
  - Finite granularity implies limits on accuracy and precision based on representation.
  - FSMs are syntactic in nature so semantics is driven entirely by context.

Epistemology 3

- There are fundamental limits on what can be done.
  - Computational complexity is like the speed of light in DFE examination.
  - DFE can never directly speak to the physical world except in limiting what FSMs can do.
  - At the edge between digital and physical systems there are assumptions.
- Benny Hill: "When you assume, you make an ASS of U and ME."
  - Be careful what you say, you may be making assumptions that are provably wrong.

Theory

- Scientific theories are not casual theories.
  - They are constructs that are testable by nature.
  - Refutation can destroy any theory, but confirmation cannot prove it.
  - Scientific theories change slowly, and normally, once accepted, only change because of dramatic changes in underlying understanding of physics, and those changes are normally only related to special or rarely seen cases.
  - Theories are different than hypotheses, which come up all the time, on a case-by-case basis.

DFE theories

- Theories in DFE examination.
  - Form a physics of information.
  - Many of them are based on mathematical results that have long been widely accepted.
  - Some of them are still conjectures, that may be proven or maybe disproven with time.
- Most such theories stem from computer engineering, computer science, finite mathematics, and related fields.
- Many such theories lead substantially limit what can be truly stated about DFE.

A standard model

- We generally interpret theory in terms of a model -- I will call it “the standard model”
  - But it's hardly standard at this point in time
- The standard model assumes laws, a judicial system with various standards
  - These are called “the legal environment” (L,R,V)
- Claims made by parties, documents, statements, and a wide variety of other non-digital information, and hypotheses are made by examiners
  - These are called “events” (E)

The standard model 2

- There is a wide variety of digital forensic evidence, typically in the form of sequences of bits
  - These are called “traces” (T)
- The DFE examiner identifies consistencies and inconsistencies
  - Between and within traces (TxT)
  - Between traces and events (TxE)
- To do this, the examiner uses forensic methods
  - These are called “procedures” (P)

The standard model 3

- Examiners work within constraints
  - There are limits on available resources (R)
  - There is an ever changing schedule (S)
- There are various implications of this model
  - The sizes of the model components
  - Available computing power and its implication on thoroughness
  - Limitations due to resources and schedule
  - Limits of currently available procedures
  - Legal limitations on what can be used, how, when, and probative versus prejudicial value

Scientific methodology in DFE

- The fundamental theorem of DFE examination:
  - **What is inconsistent is not true**
- DFE examination consists of testing hypotheses to try to refute them.
  - No matter how many tests are performed, except for special cases, ***you can't prove that anything is true***.
  - The ***best*** you can do, is show that your ***tests failed to refute*** the ***hypotheses*** at issue.
  - The ***most*** you can say (in proof) is that the ***results*** of the tests you did were ***consistent with*** some set of ***hypotheses***.

Refutation is king

- On the other hand...
  - One refutation disproves a hypothesis.
  - The ***least*** you can say based on refutation is that the ***hypothesis is not true***.
- Thus the methodology consists of:
  - Devise testable hypotheses (A ***consistent*** with B)
  - Test those hypotheses against the evidence
    - A scientific test should seek to refute a hypothesis and not to confirm it (seek ***inconsistency***)
  - Inductive and deductive logic are valuable tools for testing hypotheses
  - As is experimental technique

The experimental basis is limited

- As an area of science, DFE has a relatively small number of peer reviewed and repeated scientific experiments.
  - The total corpus is <500 serious papers.
  - Most of these have very limited applicability.
  - Most not focused on fundamental understanding.
  - Most experiments don't meet the standards of scientific rigor typically of other fields.
  - Most experiments are oriented toward confirmation rather than refutation, which makes them scientifically dubious at best.

Experiments and tools

- DFE is latent, therefore
  - Experiments require tools
- Experiments are limited by the tools, therefore
  - We need to understand the limits of the tools to understand the limits of the experiments.
- We need a methodology to evaluate tools
  - Without a methodology, regardless of what the tools tell us, we don't know how to interpret it.
- What's involved in this methodology?

Tools must be...

- We must understand the nature of errors made by tools.
  - To do this, we need an error model.
- We must understand how to calibrate tools, how to test tools, and create a systematic approach to doing so.
  - The calibration process typically involves validation with known samples.
  - The testing process typically involves verification of the software, which normally involves mathematical proofs combined with tests that exploit the error models.

Tool interpretation

- Regardless of how "good" the tool is:
  - It must be properly used
  - The results must be meaningfully interpreted
  - The limits of the tools must be understood
- This implies expertise by the examiner:
  - Knowledge
  - Skills
  - Experience
  - Training
  - Education

Note the need for a theory of measurement and its application in the context of tool usage...

- What does the ruler measure?
- Do I need the same ruler to test it?
- Can I use the same ruler to test it?
- Can I use a tool that doesn't reveal the mechanisms producing its outputs?

Information physics examples

- Digital space converges with time
  - FSM: (I,O,S,m:{IxS}→{O,S'}) IF |I|>(|O|+|S|) THEN ∃(i,i')∈I:∃(o)∈O,∃(s)∈S, i→(o,s) and I'→(o,s)
- Time is a partial ordering
  - FSM outputs are strictly sequential as sets but...
    - Traces as recorded are subject to Δt
    - When multiple FSMs are present, A≈B may apply
    - Trace time stamps subject to delays, etc.
- Time directional asymmetry
  - Given {IxS}, {O,S'} are unique and known
  - But... given {O,S'}, {IxS} are non-unique

Information physics examples

- Thorough examination, in a sense of looking at all possibilities, is almost never feasible
  - 100 instruct. (i) @ 10^9 i/s → $10^{10^{18}}$ 1s i-sequences
- Time and space are discontinuous
  - Finite granularity space (bits) and time (clocks)
- I/O not repeatable across the D-A-D interface
  - Experiment: Print a JPG, scan it back, compare
  - Experiment: Scan the same page multiple times
- Precision is always necessarily finite
  - (/ 1 3) is precise, but try for Pi

Information physics examples

- Hash functions and their limitations (lossy)
  - $|I|=2^n$, $|O|=2^m$, $n>m \rightarrow \exists i,i' \in I, o \in O \;\; i \rightarrow o, \;\; i' \rightarrow o$
  - AND $\exists o \in O, \;\; \exists \ddot{I} \subset I, \;\; |\ddot{I}| \geqslant 2^{n-m}$
- Regular physics still applies in most cases...
  - A signal traveling from San Diego to San Francisco cannot get there in less than so many milliseconds
- Computational complexity ≈ c in computers
  - After it got there, it has to perform a computation, and that also takes time!

Presentation

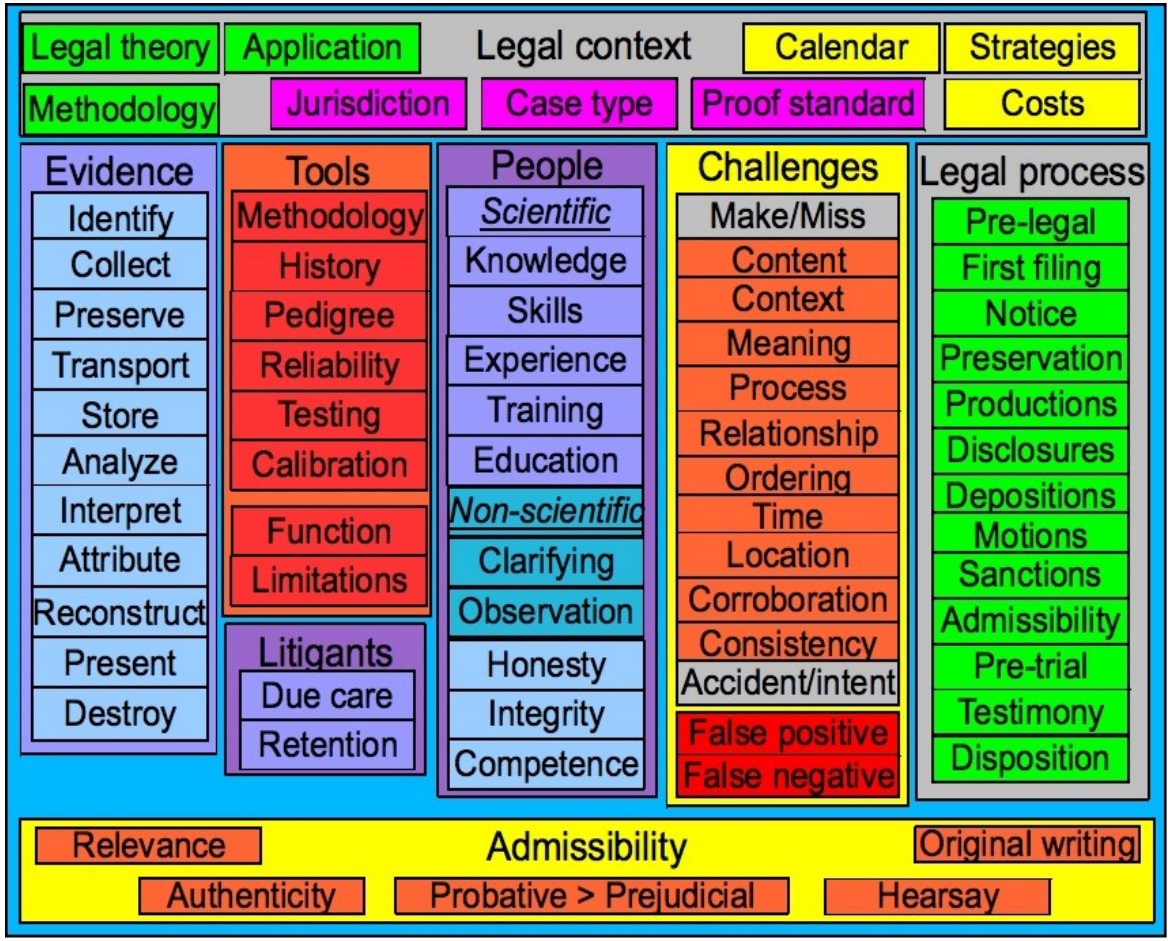
- Presentation is intimately tied to, but not directly part of, examination
  - Because DFE is latent, presentation is always necessarily an issue
    - For examining results of experiments
    - For the jury in understanding the presentation
    - For the judge in evaluating admissibility
    - For the opposition in evaluating expert reports
- Today, there is no standard for presenting the most common representations of DFE
  - Even something as simple as presenting a text file is fraught with potential errors.

Alternative presentations

- Plaintiff's sworn statements are inconsistent with the evidence.
- If Plaintiff's sworn statements are to be believed, the evidence is not.
- If the evidence is to be believed, Plaintiff's sworn statements are not.
  - The first of these statements encompasses the second two
  - The second seems to say that the evidence is lying
  - The third seems to say that the Plaintiff is lying

Outline

- Background of the speaker and subject
- Building a science
- My current approach
- Limitations
- The long view
- Your turn!



Limitations

- Some things are now widely accepted
  - The fundamental theory of digital forensics
    - What is inconsistent is not true
  - Some elements of information physics
    - You can copy without alteration, etc.
- Others are not
  - Much of information physics
    - Just starting to be taught to graduate students – subject to validation
    - Digital space converges with time (need more formal definitions) ...

Limitations

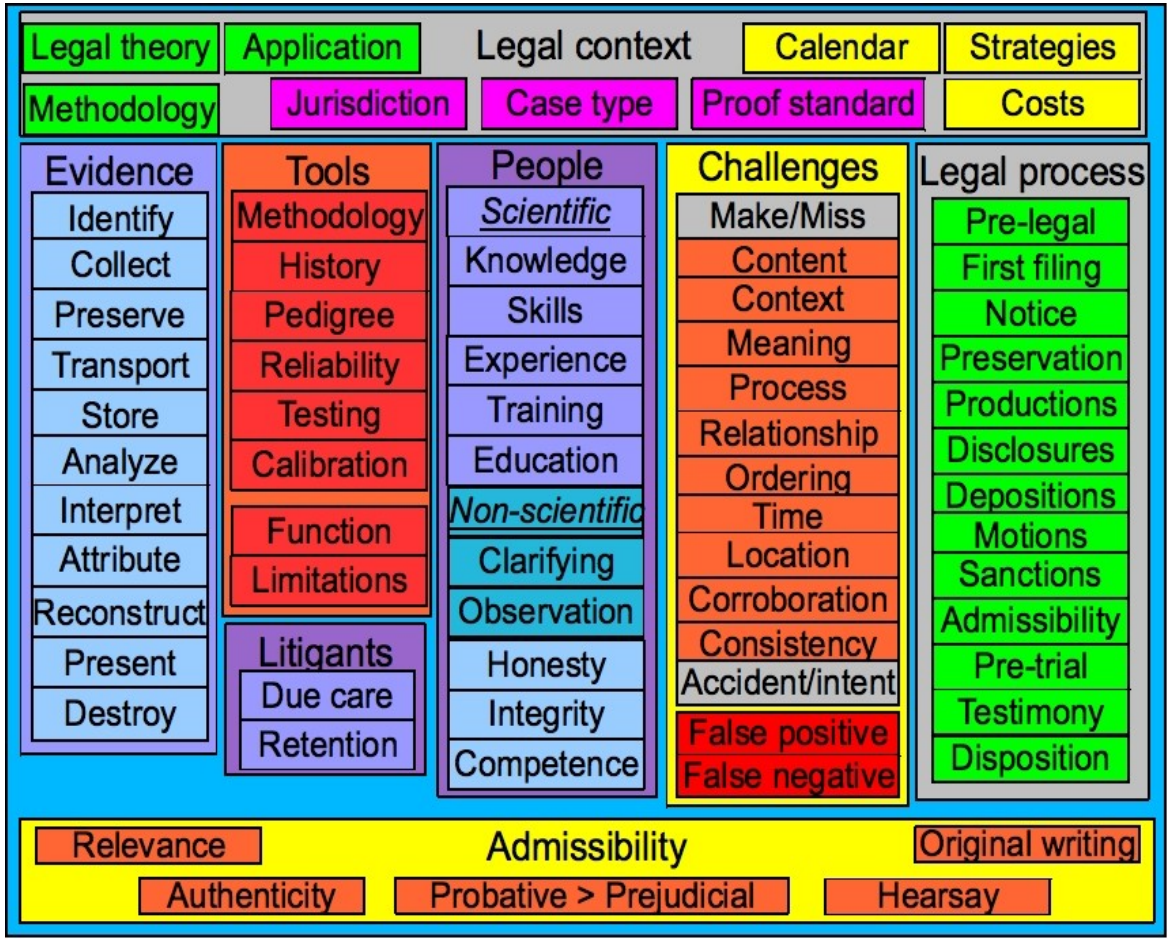
- Much of information physics is conditional
  - Like Newtonian physics and what happens as you approach the speed of light
  - Information physics is “quasi-something” …
- Methodologies for testing are problematic
  - Because of results from digital systems testing theory and issues of coverage, etc.
  - The nonlinearity and discontinuity of the digital space are problematic
  - Complete tests are to computationally complex and statistics inadequately defined …

Limitations

- I could go on all day... and I did...
  - "Challenges to Digital Forensic Evidence"
- This "science" thing is a really big issue
- And yet it is not really being addressed by the funding agencies or researchers
  - "The day is short, the task is great, the laborers are lazy, the wage is high, and the master is urging results." - Rabbi Tarfon, Mishnah Pirkei Avot
  - "A long way to go and a short time to get there" - Jerry Reed, East Bound and Down - Smokey and the Bandit

Outline

- Background of the speaker and subject
- Building a science
- My current approach
- Limitations
- The long view
- Your turn!



The long view

- The “science” problem is not unique to digital forensics
  - Information protection has many of the same problems
  - And many of the same solutions
- The thing about science is – it requires
  - Scientists (note: computer “science”?)
  - Sustained funding (vs. no funding)
  - Repetition (as in “waste”)
  - Refutation (as in “wrong answers”)

The long view

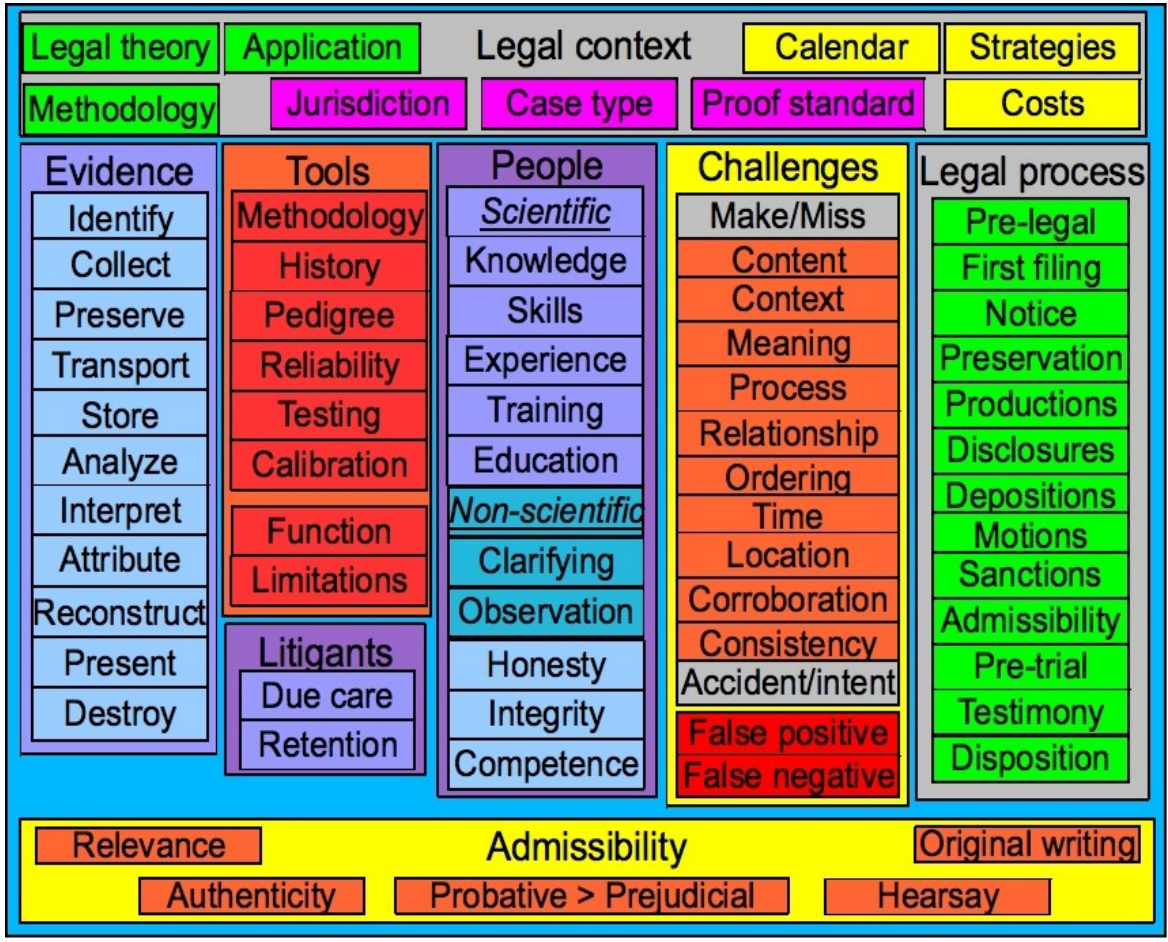
- The other thing about science
- It seems to actually work...
  - You tend to get ***more*** reliable systems over time
  - You tend to understand ***more*** about how the world works and how to make it work for you
- As opposed to what we have been doing
  - Casual theories explained with “because”
  - Evidence in the form of rumor and innuendo
  - Looking back with experience vs. forward with theory and experiment

Review of thesis

- Digital Forensic Evidence (DFE) Examination
  - Is not operating as "normal science"
    - What can we build community consensus for?
    - What well-defined and consistently used terms?
    - What well-understood epistemology?
    - What theory / methodology should we choose?
    - What strong experimental basis should we build?
    - What agreed-upon physics should we use?
- How do we move into normal science?
  - Build a community consensus!
  - Is the path I have outlined the right start?
  - What can we embrace / should we change?

Outline

- Background of the speaker and subject
- Building a science
- My current approach
- Limitations
- The long view
- Information physics
- Your turn!



Thank You

http://calsci.org/ - calsci at calsci.org
http://all.net/ - fc at all.net