THE CENTER FOR ADVANCED STUDIES
IN SCIENCE AND TECHNOLOGY POLICY

Technology as a Tool to Protect Privacy: Enterprise Architecture and Civil Liberties

K. A. Taipale
Executive Director
Center for Advanced Studies

Presented at:
"Counterterrorism and Privacy"
McCormick Tribune Foundation Conference Series
ABA Standing Committee on Law and National Security
Cantigny, IL • June 24-25, 2004

Obligatory self-promotion

Technology, Security and Privacy:
The Fear of Frankenstein, the Myth of Privacy
and the Lessons of King Ludd
Yale Law School CyberCrime Conference Paper (March 2004)

<http://www.taipale.org/papers/TSP-YLS.htm>

Data Mining and Domestic Security:
Connecting the Dots to Make Sense of Data
5 Columbia Sci. & Tech. L. Rev. 2 (December 2003)

<http://www.stlr.org/cite.cgi?volume=5&article=2>

www.advancedstudies.org

Today's presentation:
"Technology as a Tool to protect Civil Liberties"

- Technology is a *tool* and can provide neither security nor privacy
- Tools function within *technical systems*, which may or may not support certain processes or needs (e.g., security and privacy)
- Technical systems are value-sensitive *social constructions* (technology in use is never value neutral)
- Thus, it's time to talk about *enterprise architecture* that relates technology and systems architecture to business process needs (i.e., a framework that binds policy and technology)

Specifically …

- Defunding TIA was a set-back for security and a pyrrhic victory for civil liberties (classified programs, whack-a-mole, commercial interests)
- Policy cannot be developed in the abstract prior to developing technology, but neither should technology be developed without taking policy needs into account (value sensitive design)
- Long-term vision is to automate much of this through rules-based processing (code policy into intelligent agents and smart data)
- System solution is to develop a distributed architecture based on web services that supports both privacy and sharing needs

The problem: changing base conditions

- Vast data volumes and limited analytic resources requires some form of data aggregation and automated processing to better "connect the dots"
- Automated processing results in efficiencies that are challenging traditional notions of *privacy protection based on inefficiencies* (practical obscurity, checks and balances, private places)
  - Data no longer transient (always available) (*time*)
  - Data is proximate (available anywhere) (*place*)

New information economics

- The cost of *data retention* is less than the cost of *selective deletion* (search becomes more efficient than editing)
- The cost of indiscriminate *data collection* is less than the cost of *selective acquisition*
- Technical means are capital intensive not labor intensive, thus *costs per unit of information* have/will decrease (~ dataspace)
- Therefore, data largely "exists" (McNeely) and question is under what circumstances can it be *accessed*, *analyzed* and *attributed* (permissibility and "costs" of intrusion at point of mediation)

“Surveillance” economics

- High collective expectation of privacy
  - cannot watch *everybody* (~ practical obscurity by volume)
- Low individual expectation of privacy
  - can watch *anybody* (and anything?)
- Question is how we allocate the “cost” of *selective attention* (data attribution) between society/gov and the individual
- For example, compare law enforcement with intelligence
  - FBI Carnivore (analytic filter a priori) (front-load costs to gov.)
  - NSA Echelon (vacuum, analyze post hoc) (front-load costs to ind.)

Selective attention and “due process”

- Policy
  - Predicate for selective attention (preemption; efficacy=R)
  - Procedures for misuse, abuse or mistaken attention
- Technology/Systems
  - *Build-in intervention points for “due process” to function*
- Technical features to support due process
  - Distributed architecture (local control and accountability)
  - Rule-based processing (intelligent agents and “smart” data)
  - Selective revelation (anonymization and pseudonymization)
  - Authentication and audit features (“watch the watchers”)

Why "Systems"

- In the real world we do not have "security" or "privacy" instead we have a *legal and political system* that incorporates and supports both values ("due process")
- Thus, in the "data world" to achieve security and privacy we need to develop an *information management and supporting technical system* that incorporates and supports both values
- Technical systems are *social constructions* involving people, organization, procedures and technologies that enable (or constrain) certain *business processes* (e.g., security and privacy)

Thus, enterprise architecture (AE)

- Enterprise Architecture is the process of developing an enterprise-wide, integrating *framework* which incorporates:
  - *business architecture* (strategy, governance, organization, process, policy);
  - *data/information architecture* (information flows to support business processes);
  - *application (systems) architecture* (to support information flows); and
  - *technology (IT) architecture* (to enable application processes).
- Cf., OMB FEA, DodAF, DOJ LEISP, etc.

Enterprise architecture framework for CT

- Business process needs
  - Security (preempt terrorist acts)
  - Privacy (protect individual civil liberties and providers needs)
- Data/Information process needs
  - Share data (“connect the dots”)
  - Protect data (subject privacy, query, sources and methods, etc.)
- Application/systems process needs
  - “Virtual” data space (ubiquitous access and analysis)
  - Federated system, local control and accountability
- Technology needs
  - Aggregate and analyze (distributed search and collaboration)
  - Disaggregate and provide intervention points

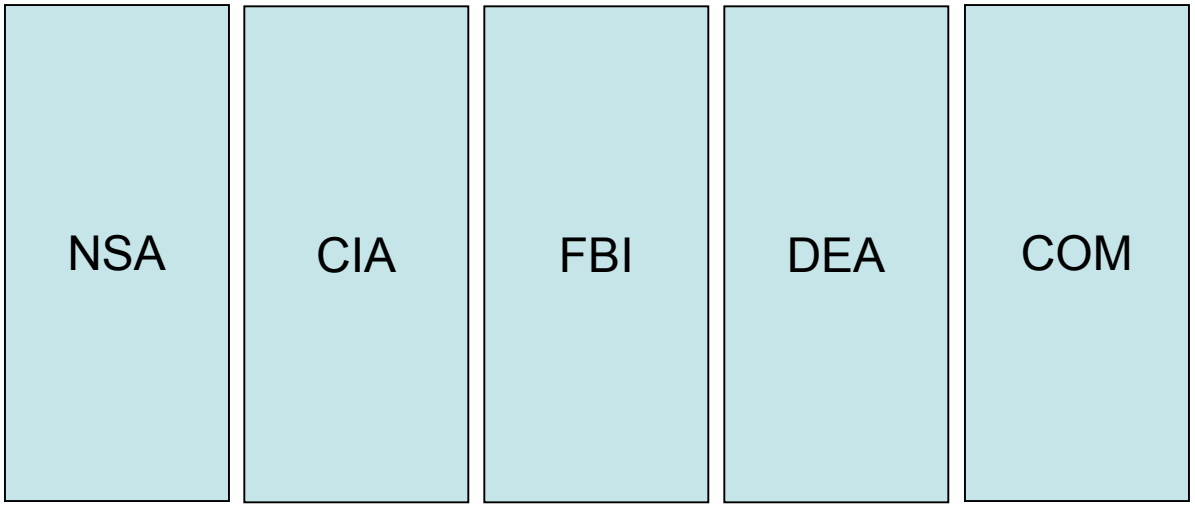
Technical Features of the CT Enterprise

- Architecture that provides/supports:
  - Distributed control and accountability *
  - Rule-based processing (~ DRM)
  - Selective Revelation
  - Authentication and Audit *

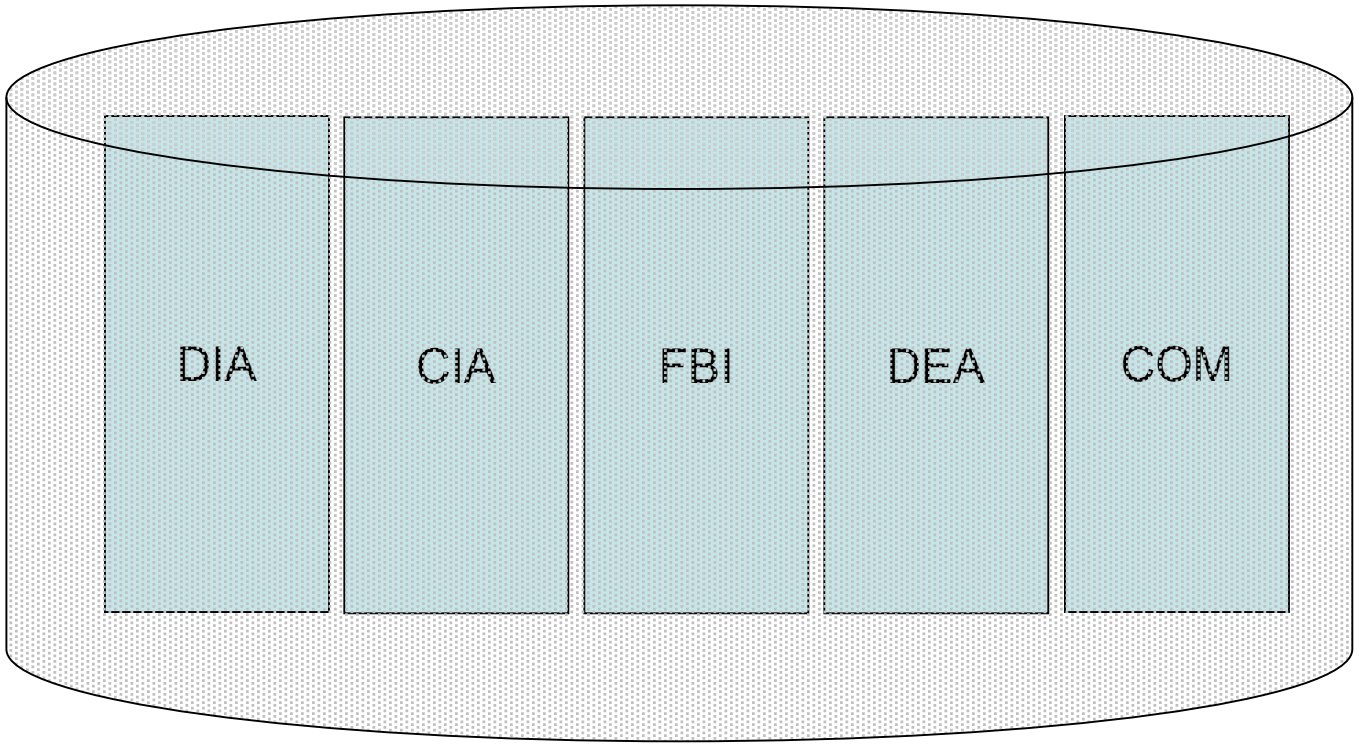
Why a distributed DB architecture?

- Efficiency (including data accuracy)
  - Pro: Collection, verifying and updating better managed locally
  - Con: Data structures support local use only
  - Technical Solution: XML (for data), web services (for providers)
- Security (including privacy of “subject of interest”)
  - P: Avoid single target for attack/abuse; retain owner control
  - C: Query (thus gov interest) exposed to multiple parties
  - T: use of indexes/directories and query privacy techniques
- Individual Privacy (including accountability for abuse)
  - P: More checks and balances and oversight with multiple owners, audit trails, and local rules
  - C: diffuse accountability
  - T: manage query/index centrally, oversight logs

Existing “stovepipe” structure



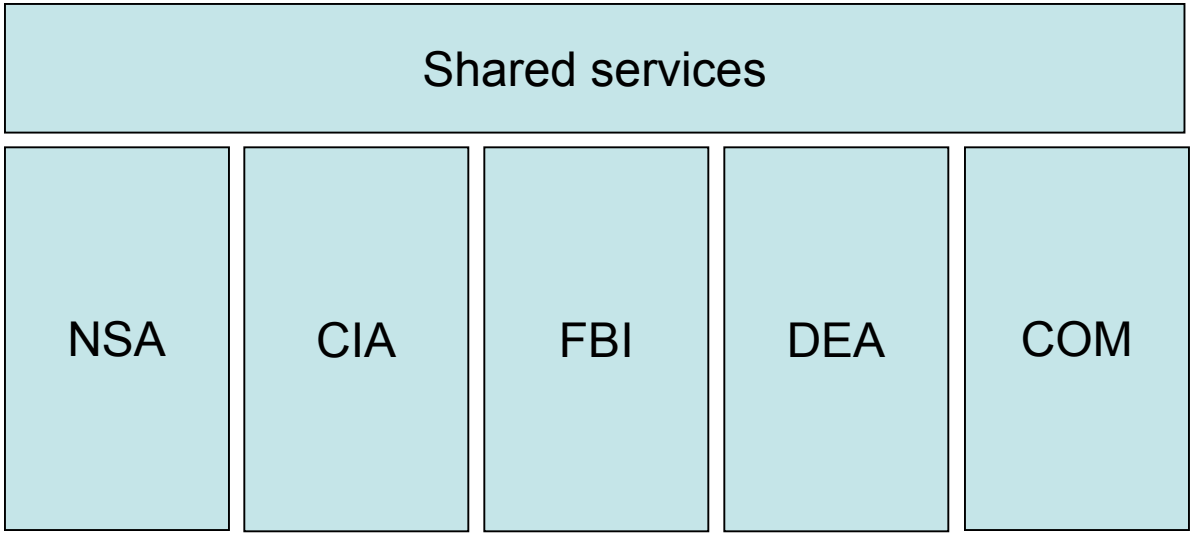
Data warehouse solution



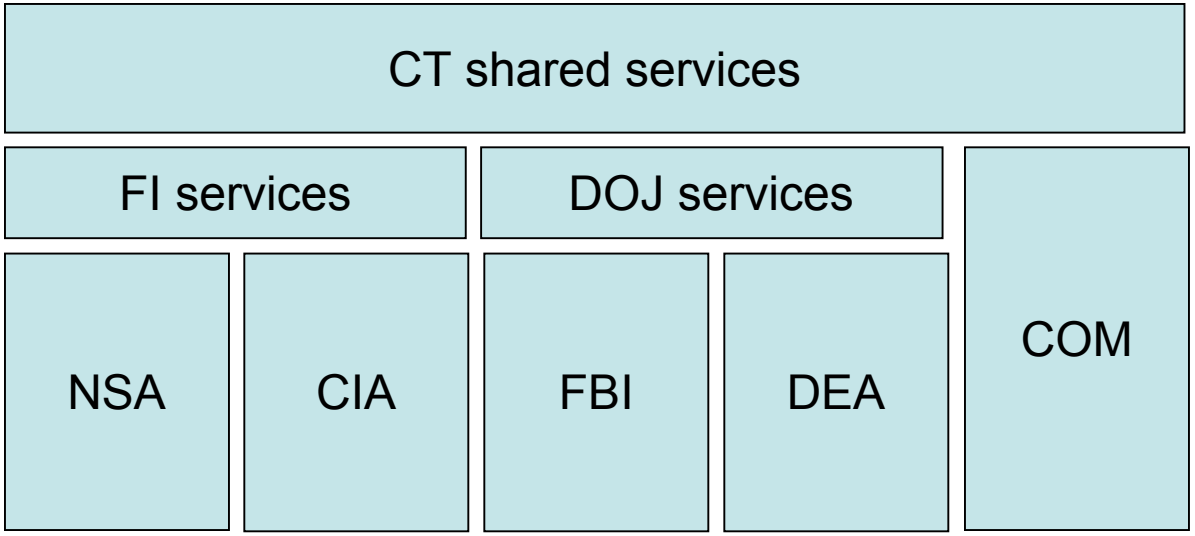
Problem is that data is not homogenous

- Originating agency “understands” its own data
  - Quality
  - Source
  - Classification
  - Privacy
  - Security
- Obstacles to sharing
  - Sharing cannot be mandated, requires local value perception
  - Forced sharing leads to concentrated (local) harm vs. diffuse (enterprise) gain (no match of incentives to business model)

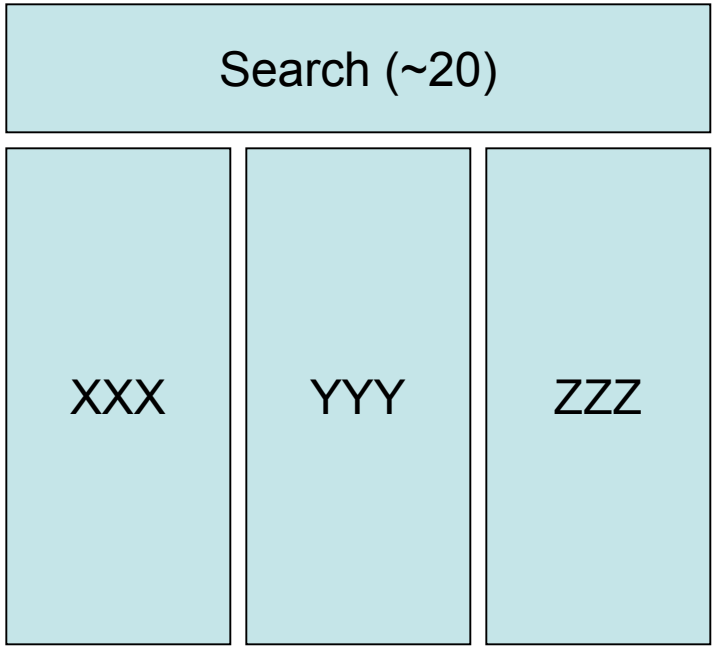
Web services solution



Layered web services solution

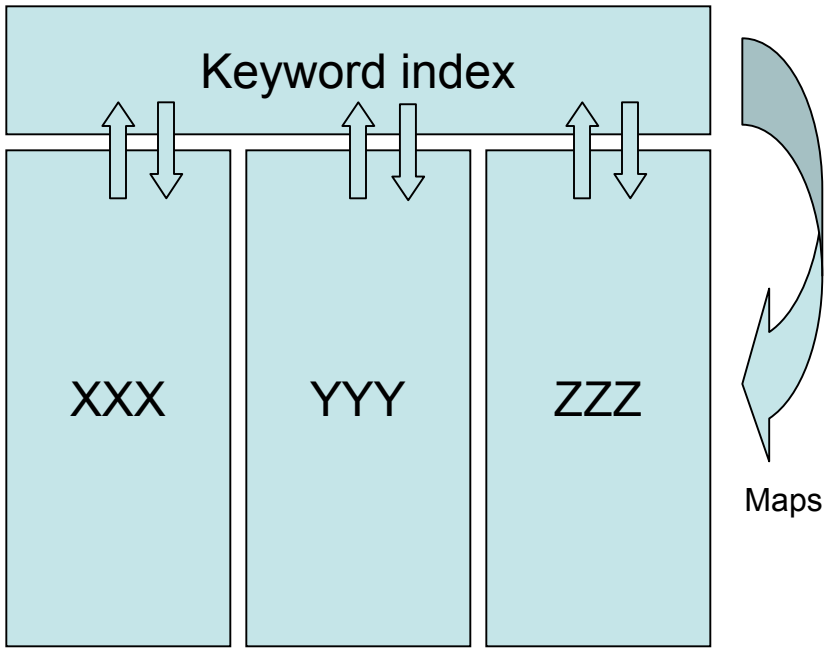


Web services example



Common search - inverted index

Existing search engines build inverted indexes that map a keyword to its precise location -- thus the index represents all the data and you have created a virtual data warehouse.



All data is recoverable from index

Directory search - access controlled DBs

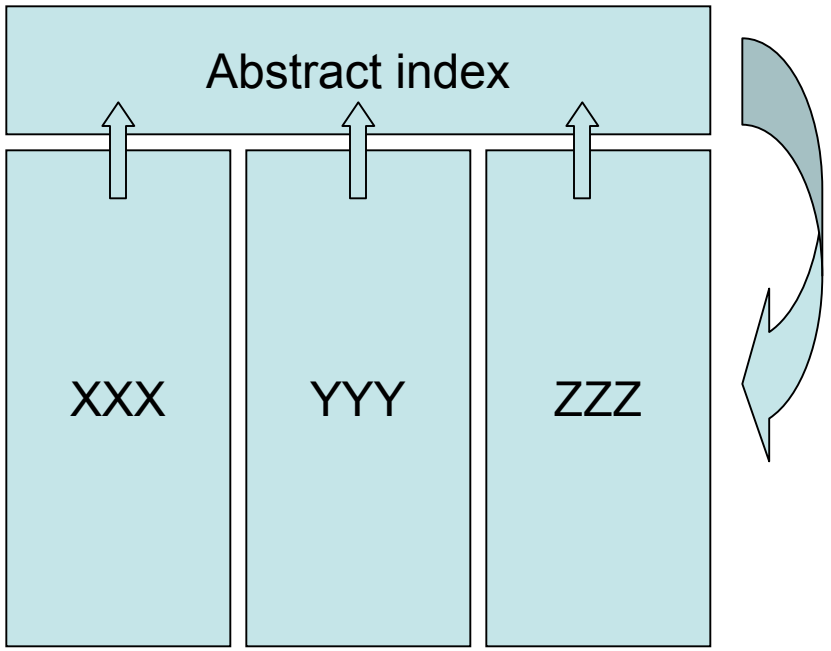
Abstract index (provider controls what data is *published* in the index – ~ automated web directories)

(~ TTIC)

(~ JTTF)

Push vs. pull

"*need to know*"



Index maps to provider, service or resource that meets local needs (originating agency continues to control access conditions)

Rule-based Processing

- Limit the scope of inquiry or use by coding policy rules
- Two aspects
  - "*intelligent agent*" - credentialed query accessing distributed DB (proof carrying code, analytic filtering)
  - "*smart-data*" – meta data labels (or wrappers) specifying information about the data and how it can be used
- ~ shrinking perimeter of defense (systems, application, data)

  (cf. "privacy appliance")

Rule-based Processing: Agent

- Intelligent Agent
  - Encrypted searches and returns
  - Negotiates access to distributed DB on local terms
  - Carries credentials authorizing access or process
- Proof carrying code
  - Local verification of agent authority
  - Local verification of agent behavior
- Difficulties
  - Efficiency
  - Scalability

Rule-based Processing: Labeling

- Meta-data labels record the relevant data attributes, e.g.:
  - Type - US person, foreign, unknown, etc.
  - Classification - unclassified, classified, secret, etc.
  - Origin - FBI, INS, AmEx, etc.
  - Reliability - rating of the source
  - Currency - date of collection or expiry
  - Governing rules - collected under EU directive, Fair Credit Reporting Act, Cable Act, HIPAA, etc.
- Labels govern subsequent processing
- Difficulties
  - requires assigning *arbitrary attributes* to data
  - derived data (data that itself results from a query)
  - legacy data (data that pre-exists labeling)
  - possible solution - research in program semantics, technologies to interpret application requirements and apply labels “on-the-fly”

Selective Revelation

- Iterative, incremental revelation of data
  - Initial revelation by statistics or categorical analysis
  - Subsequent revelation(s) justified on prior results
- Allows for data analysis without revealing personally identifying data (or other “content”, for example, classified data)
- Anonymized/de-identified data
  - Stay technologically one step ahead of trivial re-identification
- Impose appropriate legal or admin procedure before revelation
  - Make the policy/technology process “cost/benefit” appropriate to *maintain anonymity in the ordinary course of business*

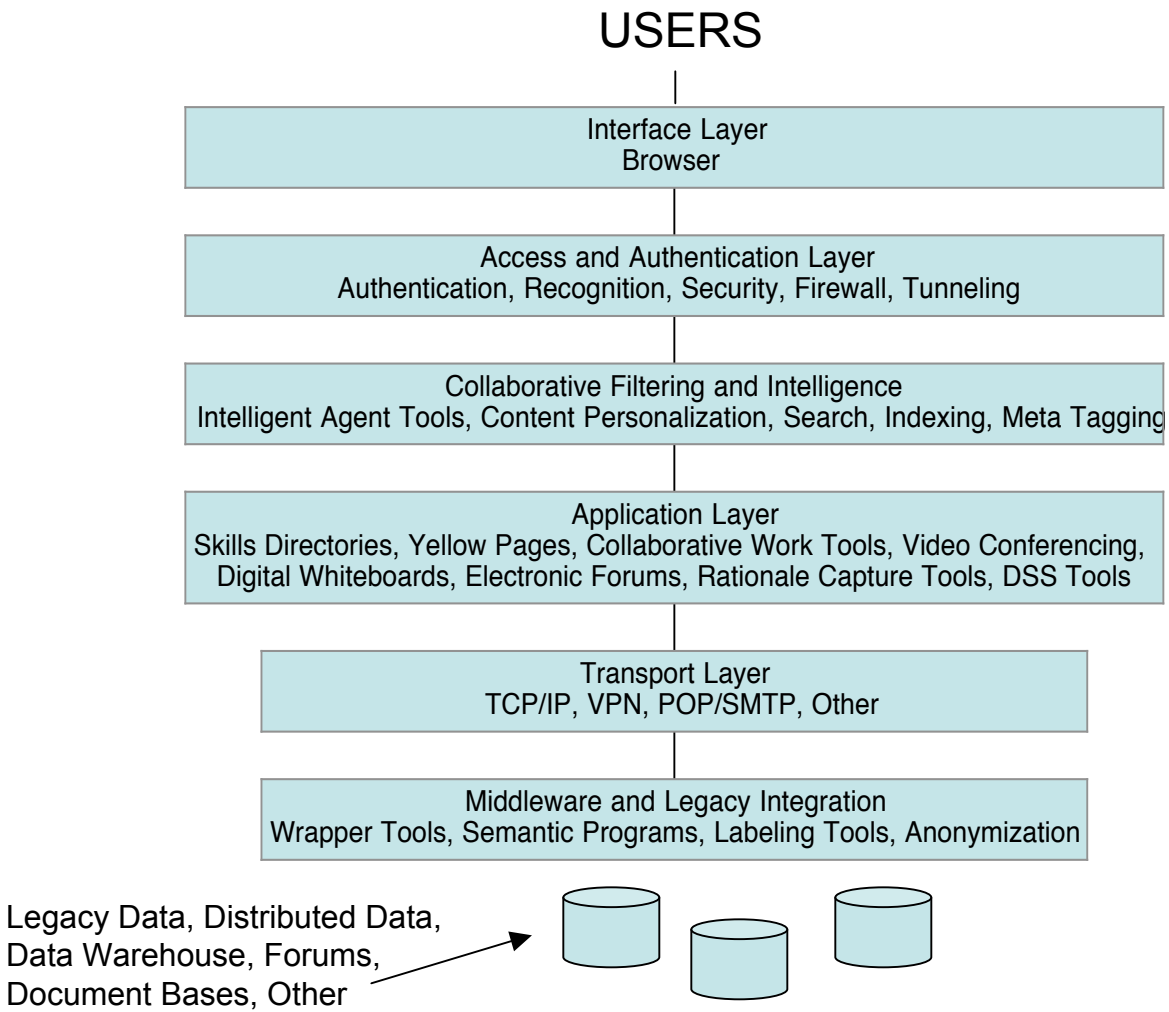
Strong Credential and Audit

- "Watch the watchers"
  - Technology creates potential for abuse but also allows for "perfect" oversight and control
  - Immutable audit trails
    - Distributed
    - Cross-organizational
    - Cross-validation
- Single biggest policy question currently being ignored:
  who controls the logs? (CIO, GC, IG, CPO, external?)
  (also, logs and PA/FOIA, etc.)
- Also, can certainty of audit substitute for predicate?

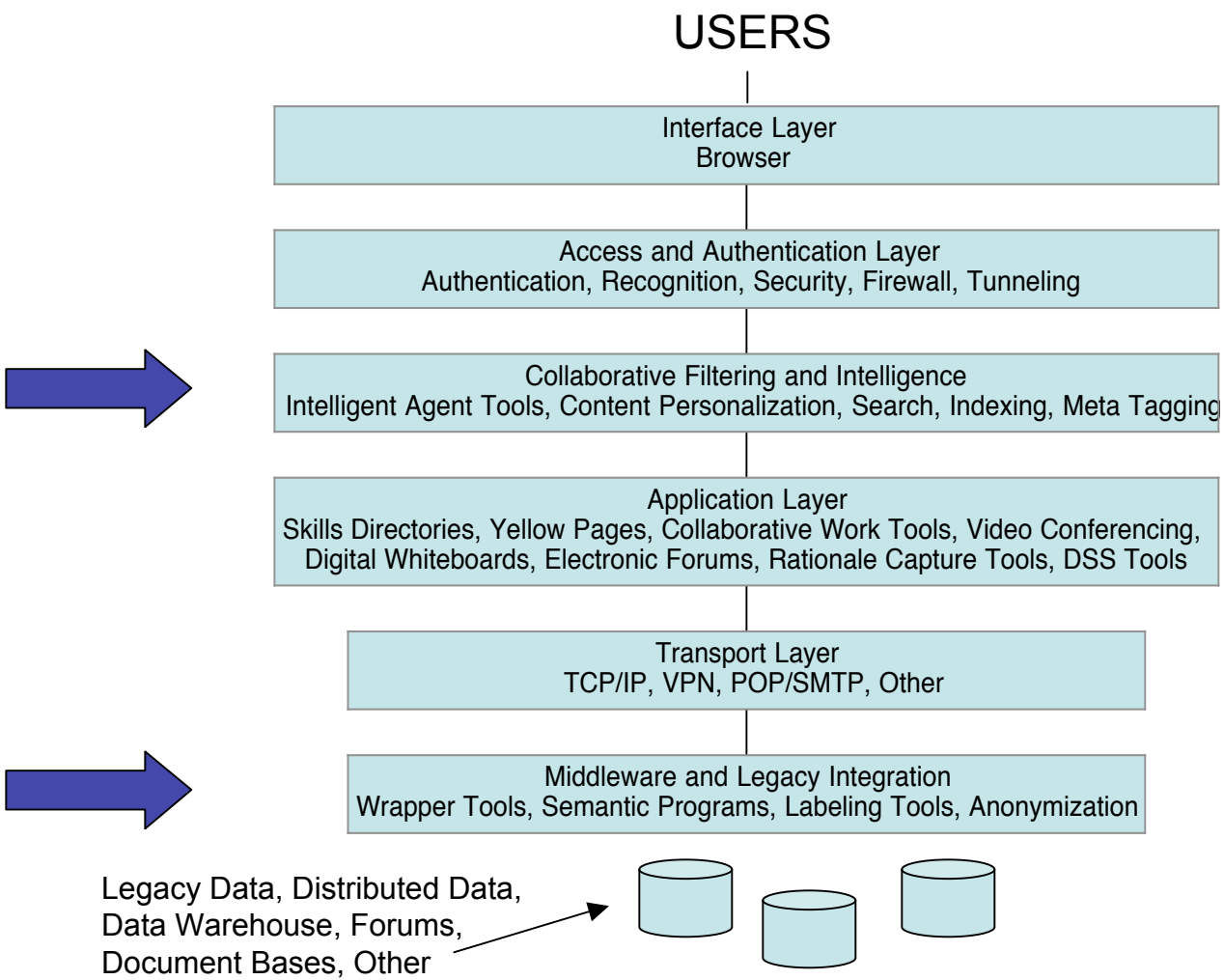
Additional Technologies

- User authentication and identification (federated ID)
  - Audit control requires user authentication and tracking
- Data identity resolution
  - Resolving data in multiple DBs with single ID or entity
- Subject identity resolution (biometrics, etc.)
  - Matching physical subject with data subject (signature eg.)
- Network security and monitoring
  - Prevent internal misuse and external abuse
- Encryption (‘controlled’ secrecy)
- Privacy and security language and protocols
  - Common protocols among systems and DBs
  - SAML, EPAL, etc., Markle IRM
- Compliance checking tools (DM logs) (turn TIA on itself)

Overall architecture and points of intervention



Overall architecture and intervention

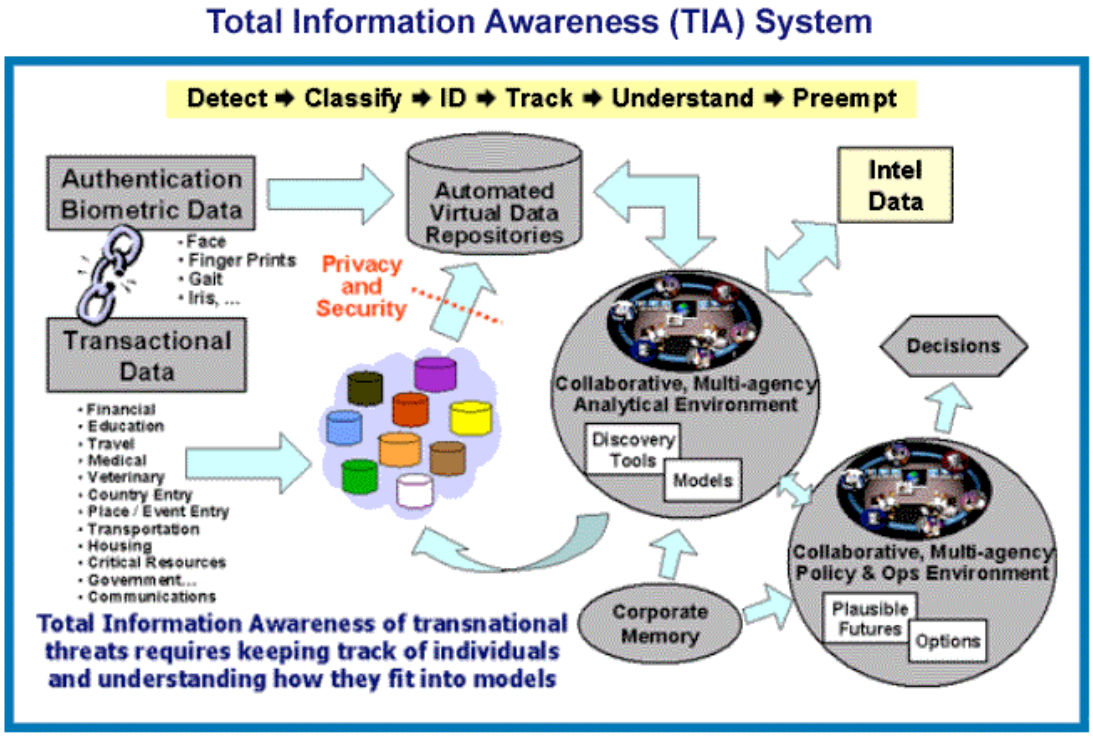


TIA/IAO: a missed opportunity

"The Information Awareness Office at DARPA is about creating technologies that would permit us have ***both security and privacy***. More than just making sure that different databases can talk to one another, we need better ways to extract information from those unified databases, and to ensure that the private information on innocent citizens is protected. ***The main point is that we need a much more systematic approach***. A variety of tools, processes and procedures will be required to deal with the problem, but they must be integrated by a systems approach built around a ***common architecture*** to be effective."

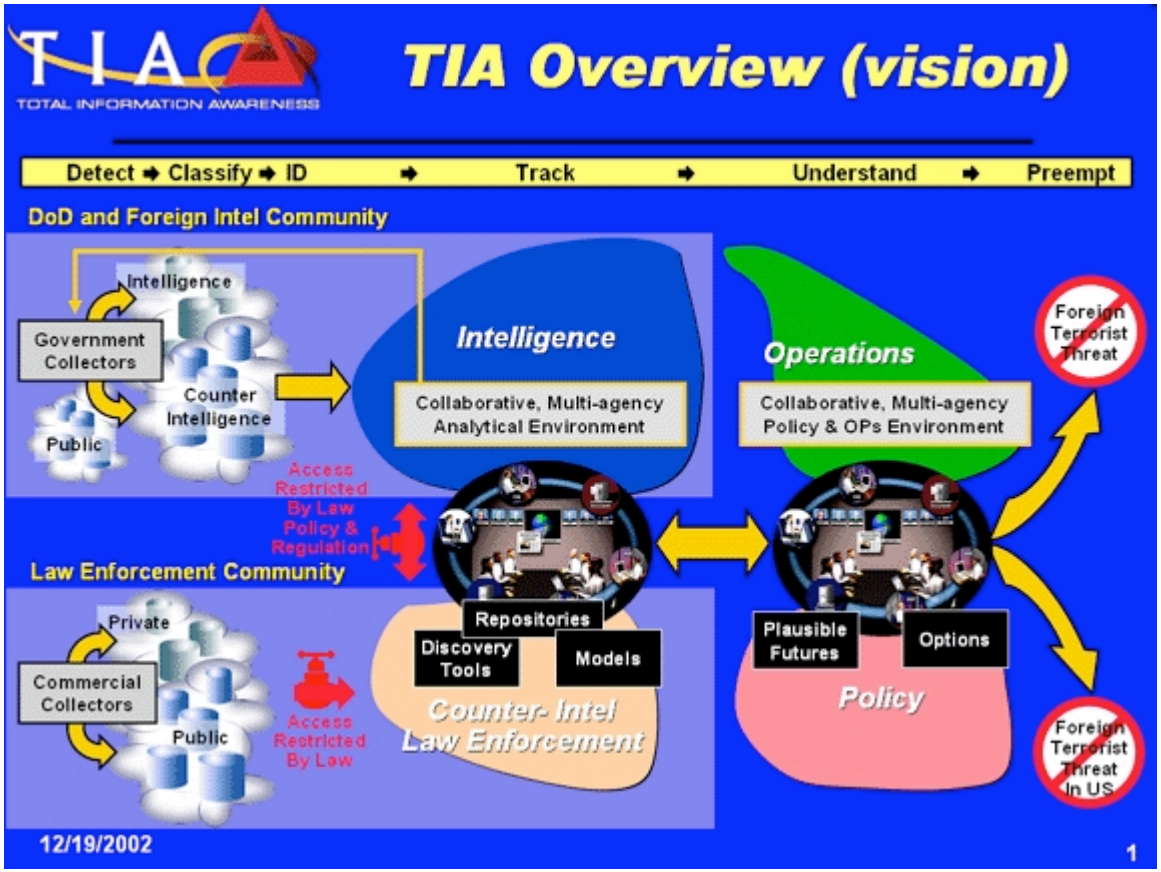
Remarks as prepared for delivery by Dr. John Poindexter, Director, Information Awareness Office of DARPA, at DARPATech 2002 Conference, Anaheim, CA, August 2, 2002

TIA



Source: DARPA's Total Information Awareness Program Homepage, http://www.darpa.mil/iao/TIASystems.htm

TIA revised


TIA
TOTAL INFORMATION AWARENESS
TIA Overview (vision)
Detect ➡ Classify ➡ ID ➡ Track ➡ Understand ➡ Preempt
DoD and Foreign Intel Community
Intelligence
Government Collectors
Counter Intelligence
Public
Intelligence
Collaborative, Multi-agency Analytical Environment
Operations
Collaborative, Multi-agency Policy & OPs Environment
Foreign Terrorist Threat
Access Restricted By Law Policy & Regulation
Law Enforcement Community
Private
Commercial Collectors
Public
Access Restricted By Law
Repositories
Discovery Tools
Models
Counter- Intel Law Enforcement
Plausible Futures
Options
Policy
Foreign Terrorist Threat In US
12/19/2002
1