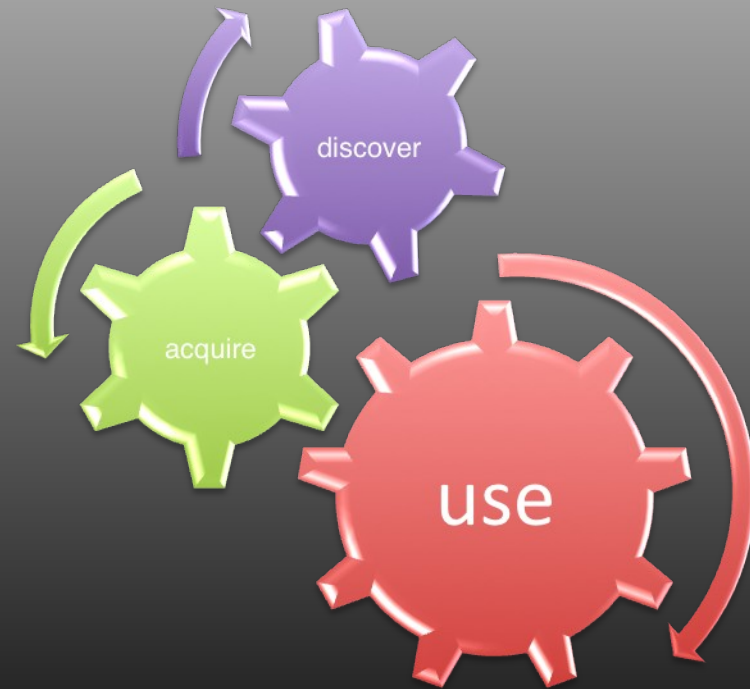


# Managing the Assured Information Sharing Lifecycle

**Tim Finin, UMBC**

08 June 2009



<http://aisl.umbc.edu/show/resource/id/498/>

# 2008 MURI project

## **University of Maryland, Baltimore County (Lead Inst.)**

T. Finin (Lead), A. Joshi, H. Kargupta, A. Sherman, Y. Yesha

## **Purdue University**

E. Bertino (Lead), N. Li, C. Clifton, E. Spafford

## **University of Texas at Dallas**

B. Thuraisingham (Lead), M. Kantarcioglu, L. Khan, A. Bensoussan, N. Berg

## **University of Illinois at Urbana Champaign**

J. Han (Lead), C. Zhai

## **University of Texas at San Antonio**

R. Sandhu (Lead), J. Massaro, S. Xu

## **University of Michigan**

L. Adamic (Lead)

*Summer  
2008  
start*

# Motivation for AIS

- **9/11** and related events illustrated problems in managing sensitive information
- Managing **Web** information & services with *appropriate security, privacy and simplicity* is increasingly important and challenging
- Autonomous **devices** (mobile phones, routers & medical equipment) need to share, too
- Moving to **EMRs** is a national goal, but raises many privacy issues
- Business needs better models for **DRM**

# Need to Know, Need to Share

- Traditional information security frameworks are based on *“need to know”*
- The 9/11 commission recommended moving from this to *“need to share”*

# Need to Know, Need to Share

- Traditional information security frameworks are based on “*need to know*”

*Unless you can prove that you have a prearranged right to access this information, you can't have it*

- The 9/11 commission recommended moving from this to “*need to share*”

*I think this information may be important for you to accomplish your mission and would like to discuss sharing it with you*

# Beyond the talking point

- There's a lot bundled into *“need to share”*
- For it to be more than a talking point, we must understand it technically, and
  - Explore its feasibility and desirability
  - Understand the ramifications, including risks and benefits
  - Develop, prototype and evaluate techniques, tools and systems to promote it

# Many underlying problems

Many barriers hinder or prevent information sharing:

- Sharing takes effort and maybe has risks. Why should I bother?
- How can I constrain how shared information is used?
- How do I know what information is available to me?
- Do I understand what the information means?
- Is the information accurate and timely?
- How can I safely let others know what I have to share?
- We're under attack and I need this information to prevent a disaster!

# Our research themes

- An information value chain of producers & consumers yields an assured information sharing **lifecycle**
- *Policies* for trust, access and use grounded in sharable **semantic models** operating in a **service oriented** architecture accelerate sharing
- New **integration** and **discovery** techniques are required to assure information **quality** and **privacy**
- Understanding and protecting the *social networks* promotes adds information diffusion and security
- **Incentives** for information sharing are required

# Assured Information Sharing Lifecycle

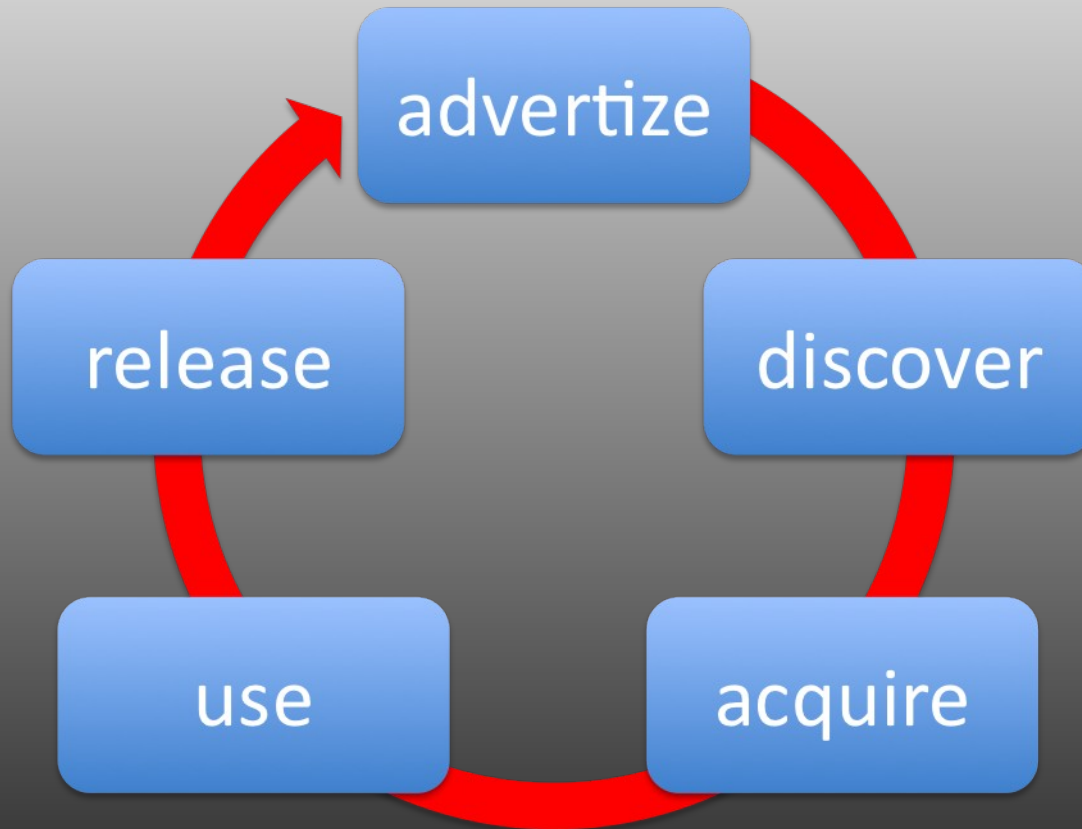
There is a lifecycle to assured information sharing that comprises information

- Advertising and discovery
- Acquisition, release and integration
- Usage and control

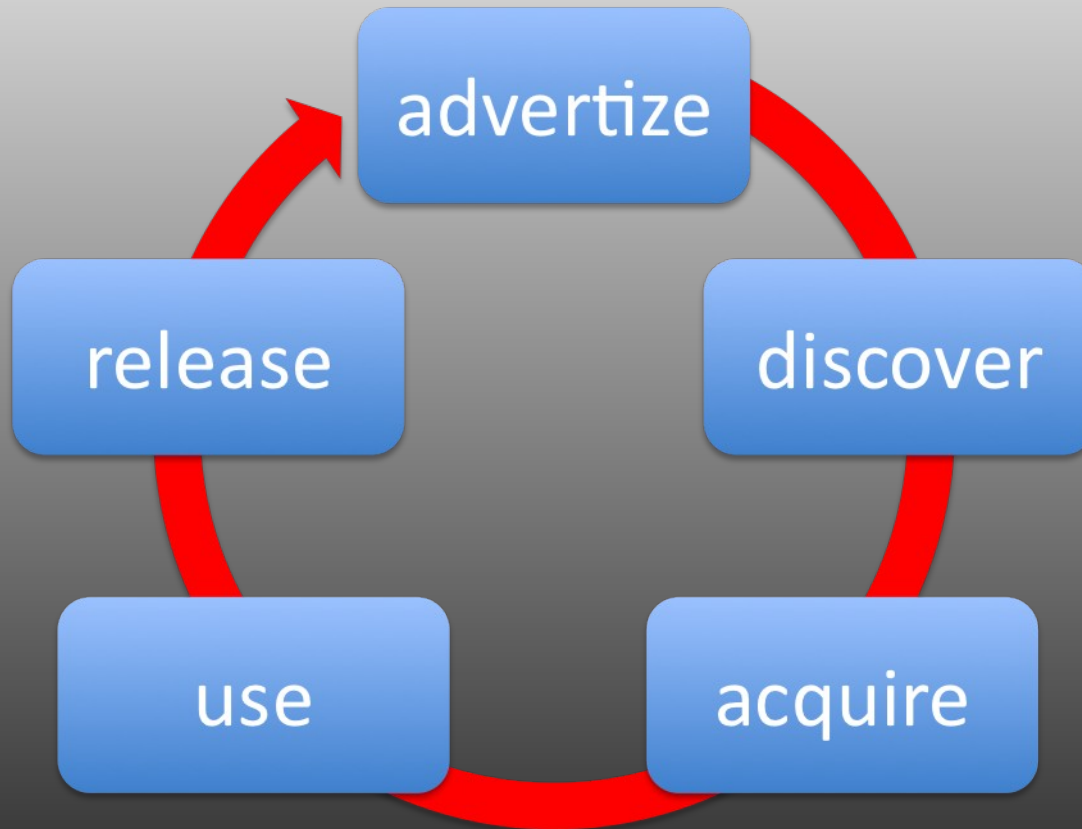
These phases realize an information sharing value chain with a network of producers and consumers



# Information value chain



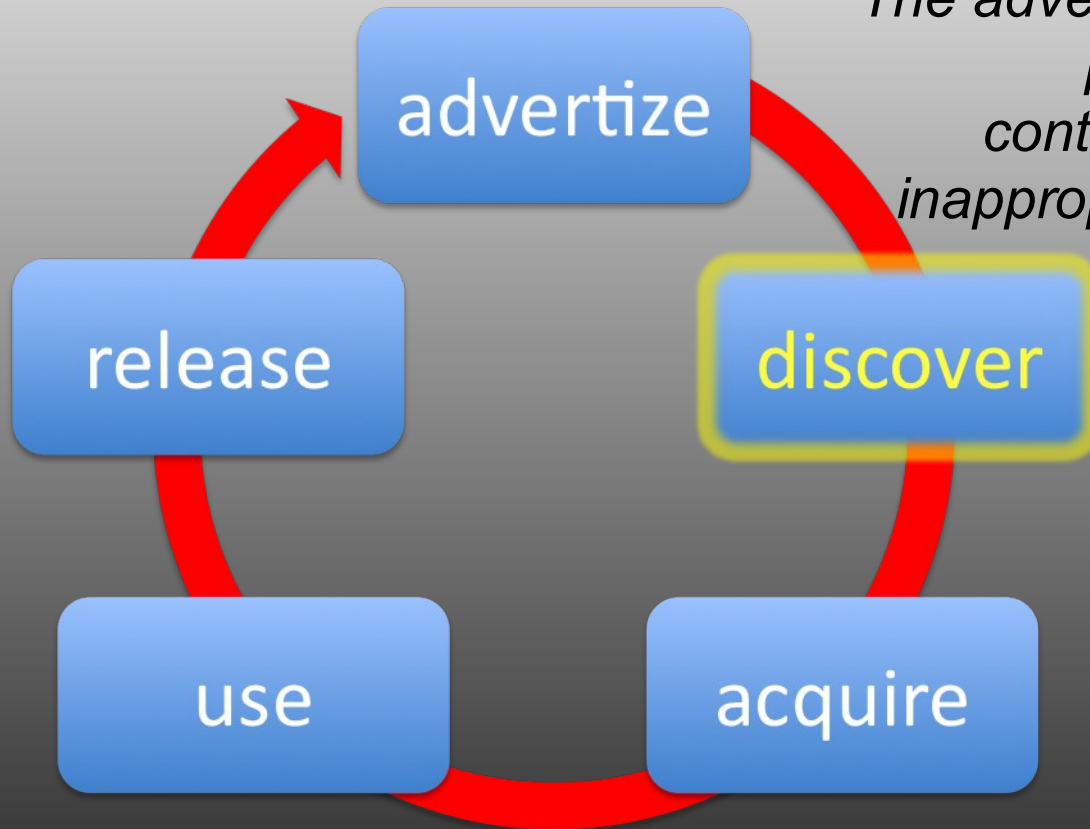
# Information value ~~chain~~ <sup>web</sup>



Potentially, everyone is both an information consumer and producer

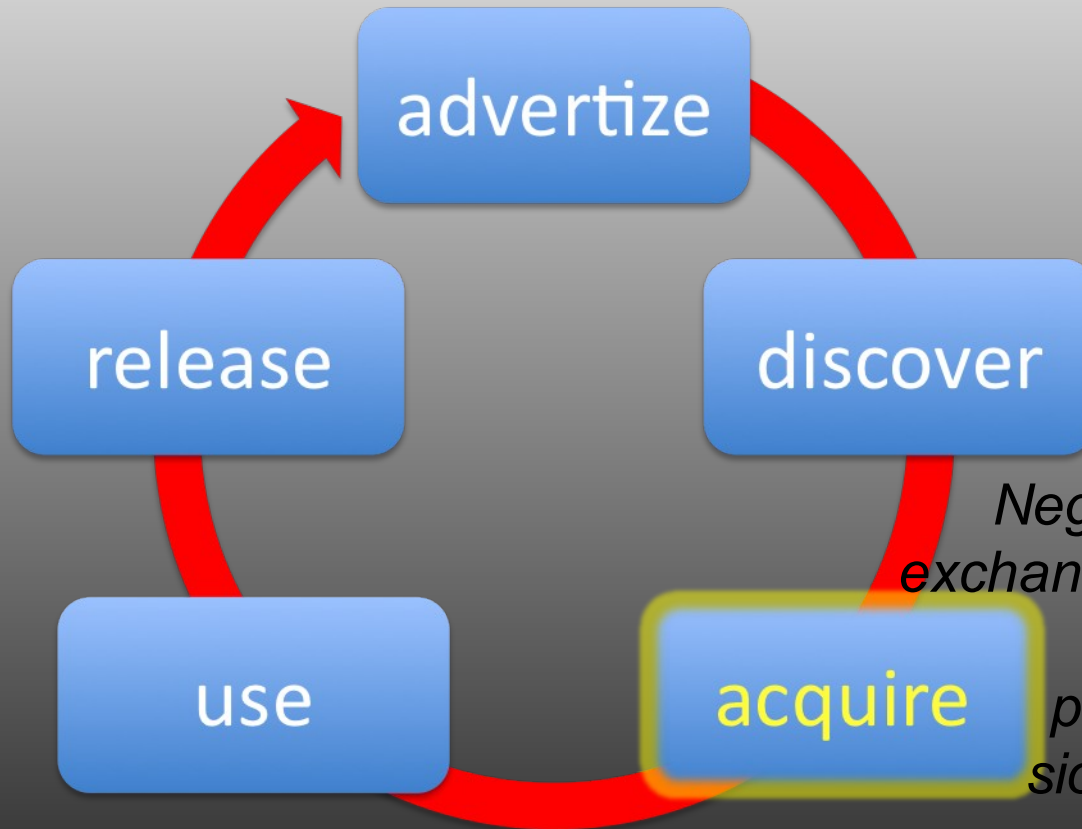
# Information value ~~chain~~ <sup>web</sup>

*The advertizing/discovery process must be controlled to prevent inappropriate disclosure*



A system discovers information it can use from the advertisements of others

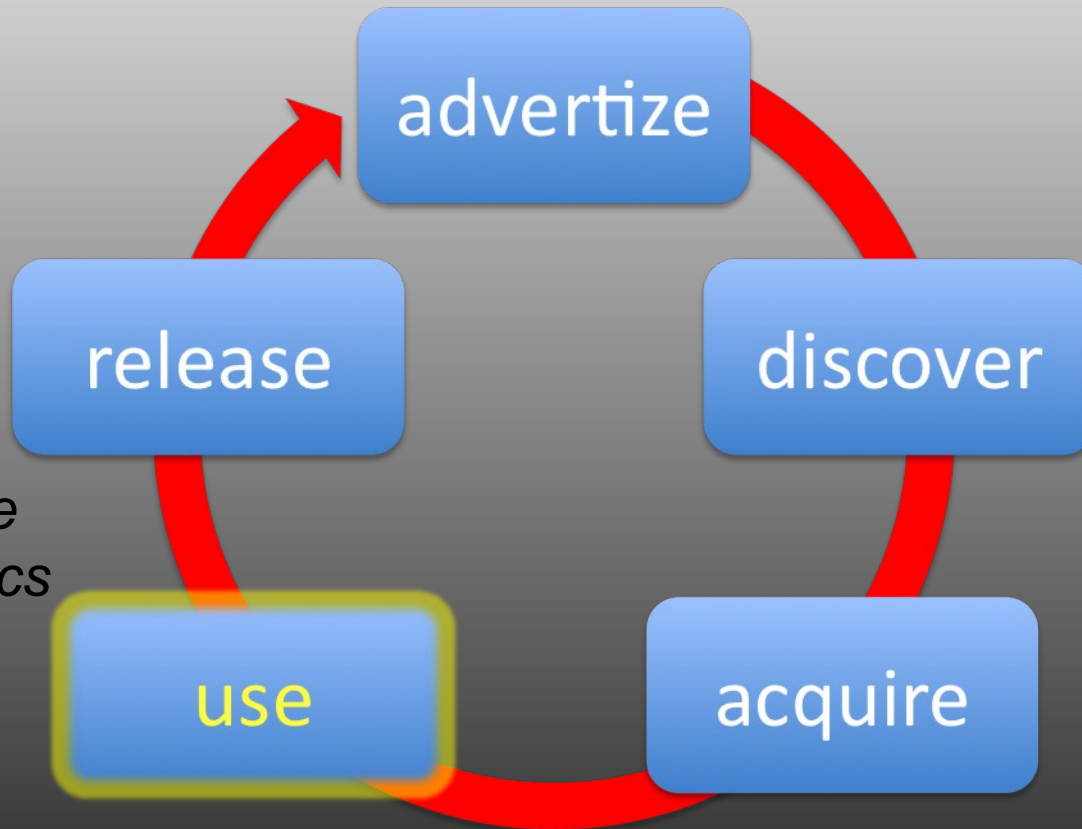
# Information value ~~chain~~ <sup>web</sup>



*Negotiation involves exchange of credentials & certificates, producing permissions & obligations*

The principles negotiate a policy for the information's acquisition and use

# Information value ~~chain~~ <sup>web</sup>

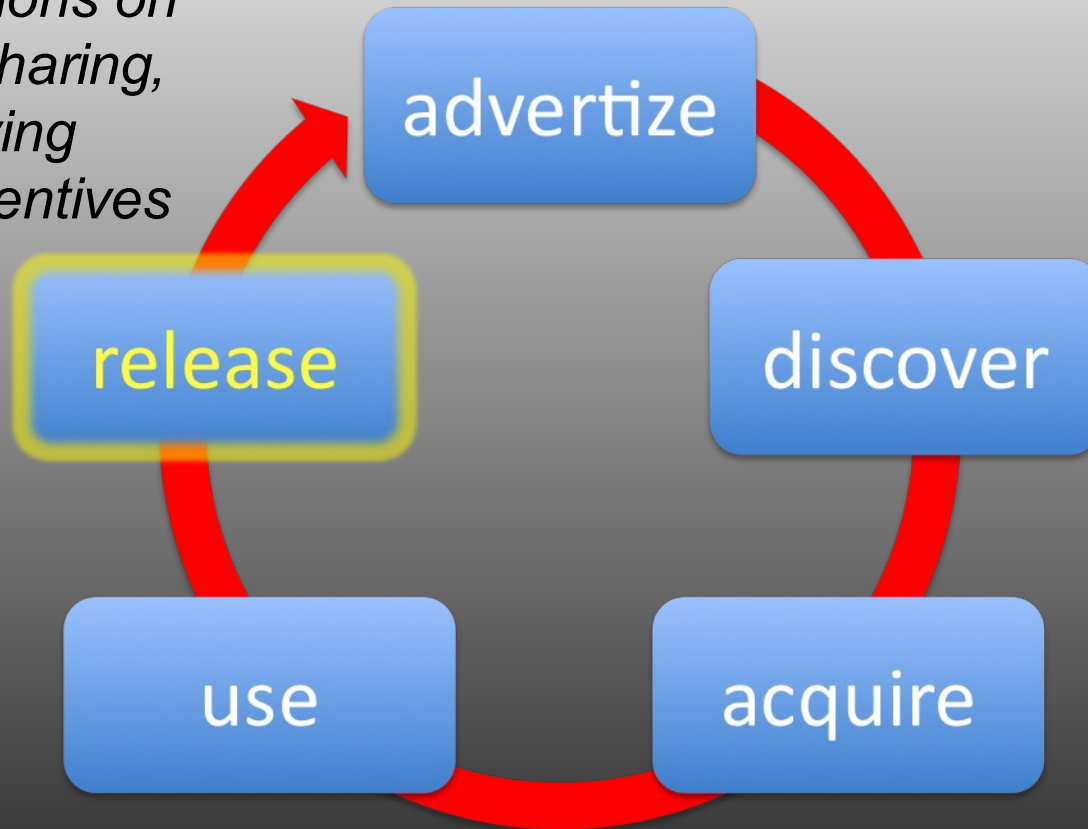


*We must assure correct semantics and information quality*

The information is used, often resulting in the discovery of new knowledge

# Information value ~~chain~~ <sup>web</sup>

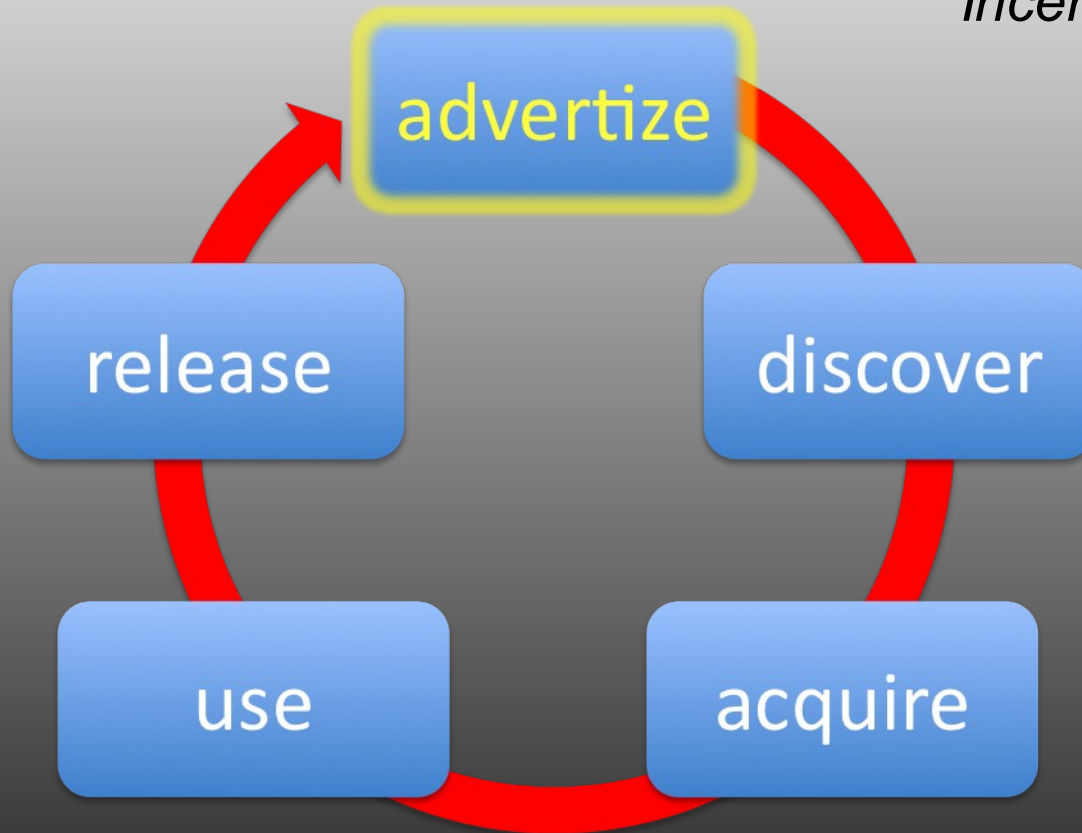
*Enforce obligations on usage and re-sharing, privacy-preserving summaries, incentives for sharing*



which is screened, adapted and summarized for possible release

# Information value ~~chain~~ <sup>web</sup>

*Incentives encourage  
offering to share  
information*



and appropriately characterized in  
advertisements for others to find

# **Our AISL research areas**

**We've organized our research into four major areas**

- New policy models, languages and tools
- Datamining, data quality and privacy preserving systems
- Social networks and incentives
- AIS service/agent oriented infrastructure

**And will evaluate our work in several integrated applications in the out years**

# Sample of AISL Recent Results

- New models, architectures, languages & mechanisms for trustworthiness-centric AIS (UTSA, Purdue)
- EXAM: environment for XACML policy analysis and management (Purdue)
- Techniques for resolving conflicting facts extracted from different resources (UIUC, Purdue)
- Study of information sharing motivation and quality in online forums (Michigan, UTD)
- Inferring access policies from logs (UMBC)
- Privacy policies in mobile/social information systems (UMBC)
- AIS infrastructure (ALL)

# But wait, there's more

- At ISI 2009 two papers from UTD
  - Ryan Layfield, Murat Kantarcioglu and Bhavani Thuraisingham, *On the Mitigation of Bioterrorism through Game Theory*, 10:15 Tuesday
  - Raymond Heatherly, Murat Kantarcioglu and Bhavani Thuraisingham, *Social Network Classification Incorporating Link Type Values*, 10:40 Wednesday
- See <http://aisl.umbc.edu/> for more

# Trustworthiness-centric AIS Framework

- **Objective:** create a trustworthiness-centric assured information sharing framework
- **Approach:** design models, architectures, languages and mechanisms to realize it
- **Key challenges, management for:**
  - *Trustworthiness* and *risk* for end-user decision making
  - *Usage*, extending simple access control
  - *Attacks*, including trustworthiness of infrastructure services
  - *Identity* extending current generation
  - *Provenance* for managing trustworthiness of data, software, and requests

# Group-Centric Secure Info Sharing

## Dissemination-Centric

- Traditional model
- Attributes & policies attached to objects (*“sticky policies”*)
- Policies enforced as objects disseminated from producer to consumer

## Group Centric

- New model
- Objects & subjects brought together as a group for sharing
- Simultaneous co-presence for access
- Two metaphors: secure meeting room; subscription service

# Progress on g-SIS

- Developed a formal model for a g-SIS system using linear temporal logic (LTL)
  - e.g., events for subjects (join, leave) and objects (add, remove), requests (read), Authz(s,o,r), ...
- Specify core properties g-SIS must satisfy
  - e.g, Simultaneity, Provenance, Persistence, Availability, ...
- Specify additional group op. properties
- Prove specifications satisfy correct authorization behavior using model checker
- See SACMAT 2009 paper

# EXAM

- The management and consolidation of a large number of policies can be an impediment to SIA
- EXAM is a prototype system for policy analysis and management, which can be used for
  - policy property analyses
  - policy similarity analysis
  - policy integration
- Focus on access control policies in XACML (Extensible Access Control Markup Language)
- Analyzer combines advantages of existing MTBDD-based and SAT-solver-based techniques

# Policy Similarity Analysis

Firefox File Edit View History Bookmarks Tools Window Help EXAM

http://192.168.0.100/EXAMFILES/htms/query4.htm

## EXAM - ENVIRONMENT FOR XACML POLICY ANALYSIS AND MANAGEMENT

### MULTIPLE POLICY EFFECT QUERY

RUN POLICY SIMILARITY FILTER

#### POLICY SIMILARITY ANALYSIS

ENTER PROJECT NAME

COMPARE POLICY  WITH POLICY

FIND ALL REQUESTS  BY POLICY POLICY9.XML AND  BY POLICY POLICY10.XML

time	[-]time				[-]memtype			[-]contype
	(21:00,22:00)	(20:00,21:00)	(22:00,23:45)	Any	weekly	monthly	video	
< 19:00	D	D	D	D	D	D	D	
(19:00,20:00)	D	D	D	D	D	D	D	
(21:00,22:00)	D				D	D	D	
(20:00,21:00)	D				D	D	D	
(22:00,23:45)	D				D	D	D	

PSA Query : Find all requests permitted by both policies.

Disjoint predicates : time cannot have two different values in any request.

Both policies permit download action when memtype is monthly and time is between 21:00 and 22:00.

Both policies permit download action to monthly subscription between 21:00 and 22:00 only if the content type is not video.

No access is permitted by both policies for video files between 20:00 and 21:00.

atch Undefined

Done

# EXAM - PSA Example

http://192.168.0.100 - Mozilla Firefox

memtype = weekly & contype = video

		[-]time				[-]memtype
	0)	(21:00,22:00)	(20:00,21:00)	(22:00,23:45)	Any	monthly
time	< 19:00					D
	(19:00,20:00)					
	(21:00,22:00)					
	(20:00,21:00)					
	(22:00,23:45)					D
	Any					

D - download

Done

		[-]time			[-]memtype		[-]contype	
		(21:00,22:00)	(20:00,21:00)	(22:00,23:45)	Any	weekly	monthly	video
time	< 19:00					D	D	D
						D	D	D
						D	D	
						D	D	
						D	D	
							D	
								D

D - download

Unconditional Conditional Non-Match Undefined

Both policies permit download of video files to monthly memberships if time is less than 19:00 or time is between 22:00 and 23:45.

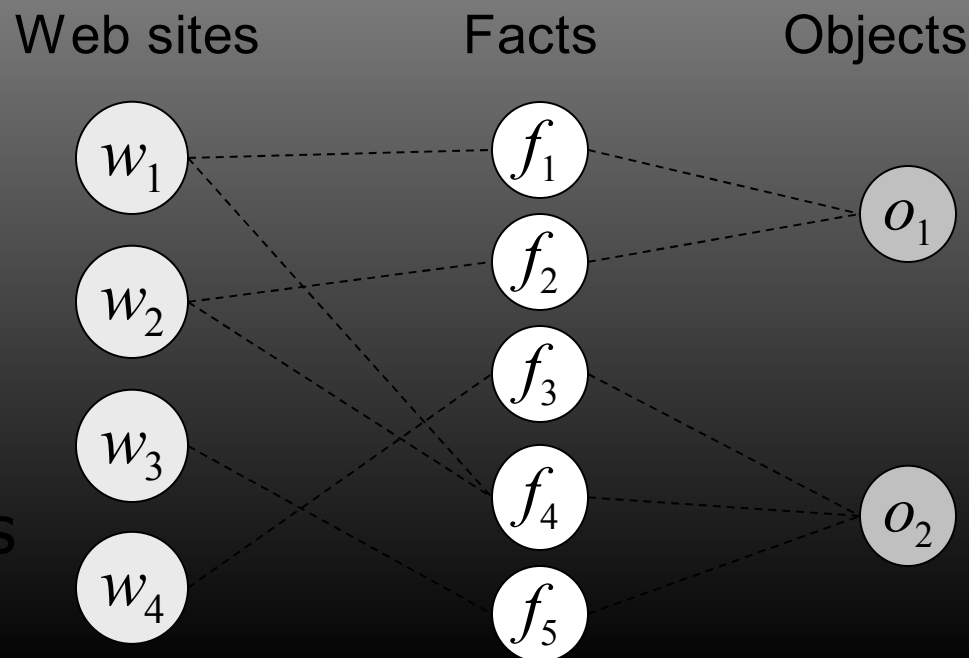
This example considers the case where membership can be both weekly and monthly.

To be demonstrated at SACMAT 2009

# Truth Discovery with Multiple Conflicting Information Providers

- **Problem:** Multiple information provider may provide **conflictive** facts on same object
  - Given different author names for a book, which is **true fact?**
- **Heuristic Rule 2:** A web site that provides mostly true facts for many objects will likely provide true facts for other objects

- **Heuristic Rule 1:** The false facts on different web sites are less likely to be the same or similar
  - False facts are often introduced by random factors



# Truth-Discovery: Framework Extension

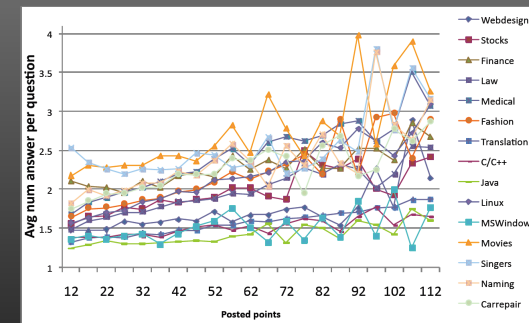
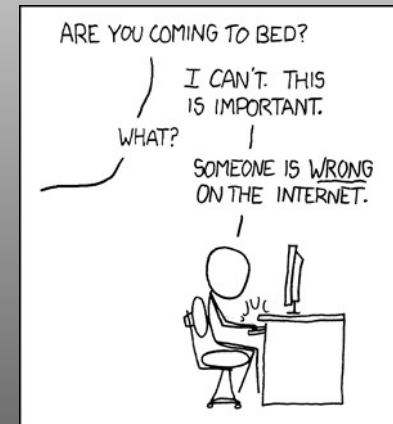
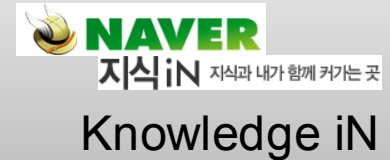
- Multi-version of truth
  - Democrats vs. republicans may have different views
- Truth may change with time
  - A player may win first but then lose
- Truth is a relative, dynamically changing judgment
  - Incremental updates with recent data in data streams
- Method: Veracity-Stream
  - Dynamic information network mining for veracity analysis in multiple data streams
- Current Testing Data Sets
  - Google News: A dynamic news feed that provides functions and facilitates searching and browsing 4,500 news sources updated continuously

# Truth-Discovery: Framework Extension

- **Multi-version of truth** *A common semantic model helps here*
  - **Democrats vs. republicans may have different views**
- **Truth may change with time**
  - **A player may win first but then lose**
- **Truth is a relative, dynamically changing judgment**
  - **Incremental updates with recent data in data streams**
- Method: Veracity-Stream
  - Dynamic information network mining for veracity analysis in multiple data streams
- Current Testing Data Sets
  - Google News: A dynamic news feed that provides functions and facilitates searching and browsing 4,500 news sources updated continuously

# Motivation & quality in information sharing

- Analyzed online Q&A forums: 2.6M questions, 4.6M answers and interviews with 26 top answerers
- Motivations to contribute include: altruism, learning, competition (via point system) and as a hobby
- Users who contribute *more often* and *less intermittently* contribute higher quality information
- Users prefer to answer unanswered questions and to respond to incorrect answers
- We can use this knowledge to design better incentive systems to encourage information sharing

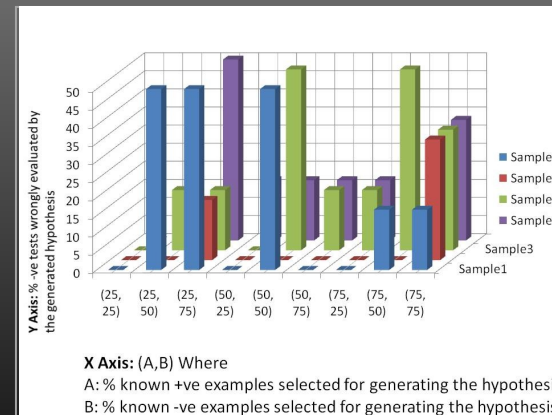
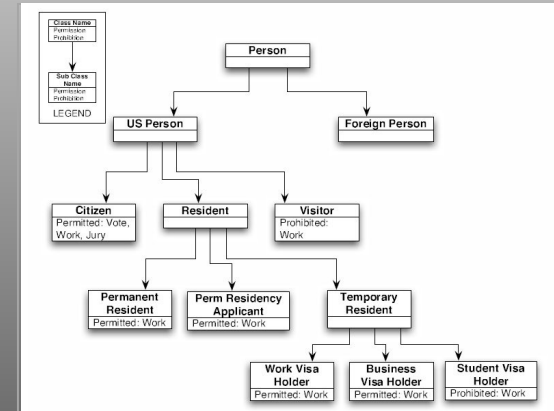


# Inferring RBAC Policies

- **Problem:** A system whose access policy is known is more vulnerable to attacks and insider threat

Attackers may infer likely policies from access observations, partial knowledge of subject attributes, and background knowledge

- **Objective:** Strengthen policies against discovery
- **Approach:** Explore techniques to propose policy theories via machine learning, including ILP and SVMs
- **Results:** promising initial results for simple Role Based Access Control policies



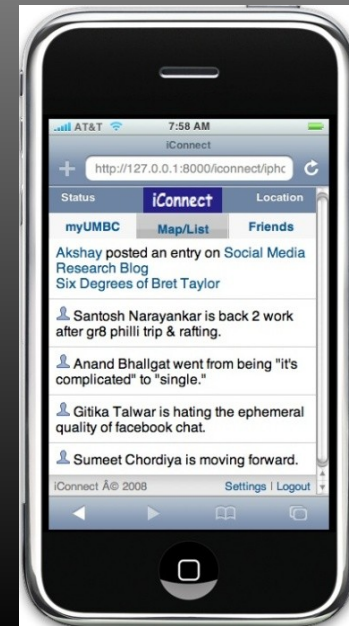
# Privacy Policies compiled to RDF N3 rules computing

- Problem: sensitive information would be shared with others
- Objective: inform users of their location activity
- Approach: iConnect a Univ
- Example: location activity

```
# Share location with teachers 9-6 weekdays
  if on campus
```

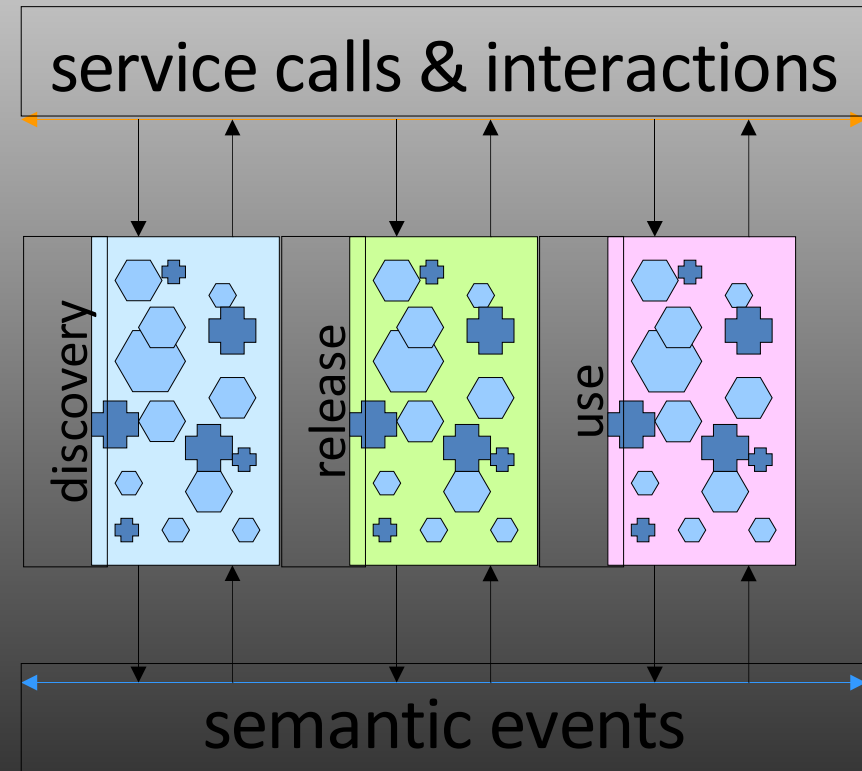
```
{ REQ a rein:Request
  REQ rein:resource LOCATION.
  ?T a TeachersGroupStuff.
  ?R a UserStuff; log:include
  { LOCATION a tu:Location; USERID a tu:Userid }.
  REQ rein:requester WHO.
  ?T a TeachersGroupStuff; log:includes
  { [] t:member [ session:login USERID ] }.
  LOCATION loc:equalTo :UMBC .
  WHO :requestTime ?time.
  "" time:localtime ?localTime.
  ?localTime time:dayOfWeek ?day.
  ?day math:notlessthan "1".
  ?day math:notgreaterthan "5".
  ?localTime time:hour ?dtime.
  ?dtime math:notlessthan "9".
  ?dtime math:notgreaterthan "18".
} => { WHO loc:can-get LOCATION }.
```

and integrate  
 ers which they  
 thers  
 system for  
 enabling end



# AIS Service Oriented Architecture

- An event-based model allows components to share **context**
- **Shared semantic models** for descriptions, communication and policies
- Initial prototype uses **Apache Axis2 SOA Framework**
- Host policy tools as services
- TODO: add enhanced agent-based protocols for advertising, negotiation and argumentation



This was just a sample of the ongoing work, see <http://aisl.umbc.edu/> for papers & more

Assured Information Sharing Lifecycle

Home User: Pass: login

**AISL**  
Assured Information Sharing Lifecycle

search AISL Search

**Assured Information Sharing LifeCycle**

Home About Us People Resources Calendar Tags Blog Wiki Subversion

AISL is a five year MURI project sponsored by the Air Force Office of Scientific Research focused on developing an integrated framework to support assured information sharing. This collaborative research project comprises researchers from UMBC, Purdue University, and the Universities of Illinois, Michigan, Texas at Dallas, and Texas at San Antonio.

Latest Resource Additions	Latest Event Additions	Latest Blog Posts
Topic Cube: Topic Modeling for OLAP on Multidimensional Text Databases	3rd International AAAI Conference on Weblogs and Social Media	How the Srizbi botnet escaped destruction to spam again
Questions in, Knowledge IN? A Study of Naver's Question Answering Community	Social Networks and Web 2.0	Scantegrity election verification system
Near-Optimal Supervised Feature Selection among Frequent Subgraphs	NSF Workshop on Data and Applications Security	NRC raises issued with datamining for counterterrorism
A Conceptual Framework for Group-Centric Secure Information Sharing	Secure Knowledge Management	Secure Knowledge Management Workshop, Dallas, 3-4 Nov 2008
D-Algebra for Composing Access Control Policy Decisions	The First ACM Workshop on AISec	Chris Clifton gives keynote at ACM workshop on AI and security

**PURDUE UNIVERSITY** **ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN** **UMBC** **UNIVERSITY OF MICHIGAN** **UT DALLAS** **UTSA**

© 2009 UMBC, Purdue University, and the Universities of Illinois, Michigan, Texas at Dallas, and Texas at San Antonio.

# Conclusions

- Assured information sharing in open, heterogeneous, distributed environments is increasingly important
- Computational policies can help
- Semantic Web technologies offer a way to share common policy concepts, policies & domain models
- Data quality and privacy-preserving techniques must be addressed
- Social aspects are important: networks, incentives
- For more information, see <http://aisl.umbc.edu/>
- Slides: <http://aisl.umbc.edu/show/resource/id/498/>