IAM Online

Persistent Identifiers for Education

Wednesday, January 12, 2011 – 3 p.m. EST
Today’s Panelists

Bill Weems, Moderator, Assistant VP for Academic Computing, The University of Texas Health Science Center at Houston

Steve Midgley, Deputy Director, Office of Education Technology, United States Department of Education

Tom Scavo, Operations Manager, InCommon

Tim Poe, Senior Collaborative Technologist, MCNC

Keith Hazelton, Senior IT Architect, University of Wisconsin-Madison

Tracy Mitrano, Director of IT Policy, Cornell University
Today’s Cyberspace Realities:

- 6,890,646,738 people on Earth
- Two billion people have Internet access!
- Facebook alone has over 500,000,000 users!
- Collaboration is no longer just among persons, but also between people and increasingly capable non-person entities (NPEs) as well as among the NPEs themselves.
- Interactions are often anonymous, but increasingly must be among identified entities that have a high degree of trust.
There needs to be a strategic discussion at the functional level of what it means for “known” entities to interact in global cyberspace when personal identity, trust, privacy and accountability must be considered.
This in turn will result in an identity management infrastructure of components, policies and procedures that allow parties interacting via the Internet to be identified and trusted on a global scale while at the same time strongly ensuring personal privacy. All this should occur with the same ease as using public Web resources.
Steve Midgley
Deputy Director
Education Technology
US Department of Education

steve.midgley@ed.gov
@officeofedtech
@stevemidgley
National Education Technology Plan Framework

Teaching

Learning

Assessment

Infrastructure

Productivity

Identity Here
Identity: Functional Requirements (non-official list)

• Satisfies FERPA (!!)
• Federated: ID’s organizations definitively and lets orgs ID other orgs or individuals
• Works without centralized ID infrastructure
• Handles anonymization w/out statistical data loss
• Handles data for operations and research
• Provides for signing as well as identity
• Is secure
• Puts power in hands of individuals to control their identity (users can set policy about how/where their identity is used)
• User-based auditing: how/where has my identity been used?
Identity: Technical requirements (non-official list)

- Easy On-Ramp: Simple implementation have simple specifications
- De-coupled stack: transport doesn’t dictate data format
- Non-proprietary / open for innovation
- Open community for development
- Extensible
- Opt-in/out participation
- Limited or no central-planning required
States and K-20

Tim Poe
Senior Collaborative Technologist
MCNC/North Carolina
Research and Education Network
tpoe@mcnc.org
In North Carolina

- 16 Public Universities
- 58 Community Colleges
- 115 Public School Districts
- 100+/- Public Charter Schools
- 48 Independent Colleges and Universities
- Independent K-12s
- Virtual Public School (2nd largest in US)
K-20 Growth

- K-20 organizations need to work together for the success of all students and maximize economic prosperity of our citizens.
- The lines between K-12, community colleges, and public and private 4-year institutions have become increasingly blurred, and we need to find ways to work together more effectively to maximize outcomes for all of our students and reduce costs.
North Carolina is developing a longitudinal data system with the intent of correlating data across K-20 to:

- Improve student outcomes
- Understand trends
- Maximize investments
- etc.

Some states/countries use 3\textsuperscript{rd} grade literacy rates to predict future incarceration rates (how many cells to build). We need to use 3\textsuperscript{rd} grade literacy rates to determine 4\textsuperscript{th} grade interventions and a lot more!!!
Mapping Across K-20

• North Carolina assigns K-12 public schools students unique identity numbers via NC WISE (centrally maintained student information system).

• Higher education institutions assign unique identities in a variety of ways.

• Intent moving forward is to use to-be-developed longitudinal data system to map various unique identities and develop data that can be used across the K-20 education.
K-12 Activities

- We have proof of concept with two K-12 school districts that are InCommon members.
- District-assigned/maintained identities are mapped to NC Wise student numbers, and students have access to a centrally maintained resource (Thinkgate) for formative assessment.
- Formative assessment will be one of the primary drivers for maturing student identity information in the coming years.
Importance of Authentication for K-12

- Many K-12 districts still do not provide authentication for K-12.
- Many K-12 resources currently do not require unique authentication.
- Scalable formative assessment will require unique authentication.
- Without unique identities for students, we have what amounts to a “50 First Dates” scenario, where resources never “know” the student.
- We need to have educational resources that are at “smart” as Netflix or iTunes.
K-12 – Moving Forward

• K-12 has an opportunity to use InCommon to leverage InCommon for access to a growing set of K-12 relevant SPs.

• K-12 needs mechanisms for effectively (and economically) implementing InCommon benefits across a variety of diverse districts. This will likely require regional or even state-wide IdPs.

• K-12 needs K-12 relevant attributes. Of particular note is a K-12 attribute that identifies grade-level. Is this an expansion of eduPerson, or eduKid? TBD.
The Grid Use Case

Tom Scavo
InCommon

https://spaces.internet2.edu/x/a4laAQ
Examples of Persistent Identifiers

• E-mail address (mail):
  – trscavo@gmail.com
  – trscavo@internet2.edu

• eduPersonPrincipalName (ePPN)
  – trscavo@internet2.edu

• eduPersonTargetedID (ePTID)
  – (trust me, you don’t want to see an example)

• Important! A persistent identifier need not be permanent (which is neither practical nor desirable).
Non-Reassignment

• A persistent identifier is non-reassigned if once assigned to an individual, it is never reassigned to a different individual.

• The non-reassignment property is required if the persistent identifier is to be used for access control
  – If a user, based on their persistent identifier, is granted access to a resource, and then later that identifier is assigned to another user…well, I think you see the problem 😊

• Fact: eduPersonTargetedID is defined to be non-reassigned whereas eduPersonPrincipalName is not
ePPN vs ePTID

- `eduPersonPrincipalName` weakens privacy by its easy associativity and by the fact that all SPs receive the same value.
- `eduPersonTargetedID` is non-correlatable since each SP receives its own unique value.
- `eduPersonTargetedID` is difficult to understand, implement and deploy, and therefore it is not widely supported by IdPs (at least in the U.S.).
Two Use Cases

• “The Grid Use Case” is actually a pair of use cases:
  1. Go TeraGrid (https://go.teragrid.org/)
  2. CILogon (http://www.cilogon.org/)

• Both grid use cases require a **persistent, non-reassigned identifier**
  – either ePTID or non-reassigned ePPN is acceptable
  – a reassigned ePPN is also acceptable if the reassignment interval is known
Use Case: Go TeraGrid

• *Go TeraGrid* allows TeraGrid users to use their campus identity to access TeraGrid resources

• *Go TeraGrid* attribute requirements:
  – persistent, non-reassigned identifier
  – See: [https://go.teragrid.org/idp.php](https://go.teragrid.org/idp.php)

• *Go TeraGrid* links the user’s campus identity to their existing TeraGrid identity
Go TeraGrid Results

• Go TeraGrid trusts 31 InCommon identity providers:
  – 12 IdPs provide ePTIDs
  – 14 IdPs provide ePPNs that are non-reassigned
  – 5 IdPs provide ePPNs that are reassigned in some circumstances

• To accommodate reassigned ePPNs, Go TeraGrid monitors account activity and requires the user to repeat the account linking process if the inactivity period exceeds the reassignment interval
Use Case: CILogon

The CILogon project facilitates secure access to X.509 certificate-based CyberInfrastructure (CI)

CILogon attribute requirements:

- persistent, non-reassigned identifier
- first name (givenName) and last name (surName or sn)
- displayName will be accepted in lieu of givenName and sn
- e-mail address (mail)
- [optional] eduPersonAssurance
- See: https://wiki.cites.uiuc.edu/wiki/x/iJYXAw
Observations

• A significant proportion of InCommon IdPs assert eduPersonPrincipalName that is not reassigned (!)
  – How do we leverage this fact?
  – In the case of reassigned ePPNs, how does the IdP communicate the reassignment interval to the SP?

• The fact that ePTID is non-correlatable is irrelevant in the case of CILogon since other correlatable attributes are required.

• Are IdPs willing to deploy ePTID or is non-reassigned ePPN “good enough?”
Consent-based SSO

• The emerging consent-based SSO model:
  – SPs encode attribute requirements
  – IdPs generalize attribute release policies
  – Users consent to attribute release

• How Better Attribute Management Helps Federation
  – https://spaces.internet2.edu/x/TRWp

• uApprove (from SWITCH)
  – This functionality will be incorporated into Shib IdP v3.0
  – Demo: https://aai-demo.switch.ch/secure-uApprove/
Source: “Shibboleth and uApprove at University of Michigan” (Tracy & Hammer, June 2010)
More Observations

• User consent is a potential game changer!
• How do SPs encode “persistent, non-reassigned identifier?”
• Are IdPs more willing to release complex identity attributes (e.g., CILogon) in conjunction with consent?
• Will users release ePPN to SPs?
  – A user can consent to ePPN but there’s little hope of obtaining user consent for ePTID
Bill Weems
Assistant VP for Academic Computing
The University of Texas Health Science Center at Houston
National Strategy for Trusted Identities in Cyberspace

Creating Options for Enhanced Online Security and Privacy

June 25, 2010

Draft

http://www.dhs.gov/xlibrary/assets/ns_tic.pdf
A common understanding of the following concepts as they apply to cyberspace is required:

- **What is “Trust”?**
- **What is “Identity” in cyberspace?**
- **What is “Privacy”?**
- **What is “Identify Theft”?**
- **What is an “Authentication Credential”?**
- **What is an “Identifier”?**
Trust, in the social contest, has several connotations:

- the willingness of one party (trustor) to be vulnerable to the actions of another party (trustee);
- reasonable expectation (confidence) of the trustor that the trustee will behave in a way beneficial to the trustor;
- risk of harm to the trustor if the trustee will not behave accordingly;
- requires some shared knowledge of identities among the “trusting” parties!
What Constitutes Identity in Cyberspace?

- A person’s identity has two components:
  - *Physical Identity* - absolutely unique to each person and can be certified by a credential provider (CP) that vets and records various physical characteristics of that person.
  - *Personal attributes* – named qualities or characteristics inherent or ascribed to a physical person and often verified by trusted attribute providers.
Privacy

• A person’s desire to restrict knowledge of some of his or her personal attributes to only those individuals who can be trusted.

• Individuals entrusted with protecting “private personal” information, must be identifiable, accountable and be given the tools to appropriately execute their responsibility.
Privacy

- **Privacy is the subjective condition a person experiences when two factors are in place:**
  - First, he or she must have the power to control information about him-or herself!
  - Second, he or she must exercise that control consistent with his or her interests and values!

- **The first factor goes to the existence of choice, the legal and/or operational power to control the release of information, not how pleasant the choice is.**

[http://www.privacilla.org/fundamentals/privacydefinition.html](http://www.privacilla.org/fundamentals/privacydefinition.html)
Identity Theft

- Someone pretending to be you in cyberspace and being trusted as you by relying parties.
- It is a form of fraud or cheating of another person's physical identity in which someone pretends to be someone else by assuming that person's identity, typically in order to access resources or obtain credit and other benefits in that person's name.
- It is a misnomer, since it is not literally possible to steal an identity as such – a more accurate term is impersonation.
Certified Authentication Credential May Be Presented Initially to Any Relying Party
Your authentication credential when presented to relying parties:

1. can only be activated by you,
2. positively identifies the certifying authority (CA) that is attesting to your physical identity,
3. asserts a defined level of assurance (LOA) that the credential is presentable only by you.
4. positively identifies you to relying parties by providing your subject unique identifier, and
5. depending on the CA, may or may not provide personal attributes other than the physical identifier.
Using a Permanent Subject Unique Identifier

Sources of Authority
- e.g. Registrars, IRBs
- Attribute Management Policies, Procedures

Personal Attributes
- UE8254K@staru.edu

Personal Attribute Provider (AP)
Organization C

Attribute Request

Service Provider (SP)
Organization B

Attribute Request

Personal Attribute Provider (AP)
Organization D

Authenticated Credential
- UE8254K@staru.edu

Identified & Credentialled Person
- Subject Unique ID = UE8254K@staru.edu
- Organization A

Sources of Authority
- e.g. certifications, etc.
- Attribute Management Policies, Procedures
Value of Authentication Credentials with Public Identifiers

- Relying parties instantly "know" it is the certified physical person they are trusting!
- An individual can use a single authentication credential to establish unlimited trust relations.
- Risk of identity theft becomes very unlikely!
- A person is "in control" of his/her "trusted" actions in cyberspace.
- One has "increased control" over release of his/her personal attributes.
- Ease of use!
Appreciating the Need for Multiple Identifiers!

- **Two extremes (i.e. boundary conditions)**
  - *Public identifiers* permanently and uniquely assigned to each physical person.
    - Greatly facilitates collaboration, trust and accountability
    - A concern: can be used to compare, correlate and inappropriately release personal attributes
  - *Private identifiers* where each relying party receives a unique, opaque identifier for the same person (e.g. eduPersonTargetedID (ePTID)).
    - This reduces the risk of using ePTIDs to correlate attributes.
- **How to create and use appropriate ID systems?**
The Challenge of Person Identifiers

Keith Hazelton
University of Wisconsin-Madison
The Challenge of Person Identifiers

- There are legitimate needs to correlate information on one individual across multiple systems.
- There are valid personal, custodial and legal concerns about a single, globally unique, permanent, portable and public identifier per individual.
- Is there a solution that addresses both the needs and the concerns?
The Challenge of Person Identifiers

- A single, globally unique, permanent, portable and public identifier meets the need to correlate, but ignores the concerns.
- There are methods for selectively linking identifiers with active consent of the individual.
- This would meet both needs.
- These methods are less convenient for the developers of systems.
- Convenience should not trump valid concerns.
Panel Discussion and Questions

Bill Weems, Moderator, Assistant VP for Academic Computing, The University of Texas Health Science Center at Houston

Steve Midgley, Deputy Director, Office of Education Technology, United States Department of Education

Tom Scavo, Operations Manager, InCommon

Tim Poe, Senior Collaborative Technologist, MCNC

Keith Hazelton, Senior IT Architect, University of Wisconsin-Madison

Tracy Mitrano, Director of IT Policy, Cornell University
Survey
Please complete the survey about today’s IAM Online:

Next IAM Online www.incommon.org/iamonline
Wednesday, February 9, 2011 – 3 p.m. EST
Provisioning - Google Groups and Federated Provisioning
   Nathan Dors, University of Washington
   Tom Zeller, University of Memphis

Thank you to InCommon Affiliates for helping to make IAM Online possible.