Opportunities for Federated Identity Management in Scholarly Work

Mike Conlon
University of Florida
mconlon@ufl.edu

Who did What?

• We want to know, with some certainty, who (people) did (association) what (scholarly work)
• Multiple assertions of the same person, the same work, the same association leads to duplication of identifiers for works, people and association – thus requiring disambiguation
People – who is who

- Associating people to identifiers stored in computers
- Trust
- Assertion

Identifying People

- We give people numbers so that we can store and retrieve information about people in computers
- Names and addresses don’t work
- Names alone, certainly does not work
How sure are we of the assignment?

• Called Level of Assurance
  – Level 1: Self asserted. Not sure at all
  – Level 2: Normal business use. Transcripts, etc. Vetting, passwords
  – Level 3: Secure use. Medical. Certificates, dual factor, annual reissue
  – Level 4: NSA use. Biologics, plus above

Two People Should not have the same identifier

• Two people with the same SSN
• Two people with the same UFID
• Two people with the same last name, first name
One person may have many identifiers

• SSN
• Drivers License
• Passport number
• Institutional ID
• eraCommons identifier
• Etc.

Cross Walking identifiers is common

• Store many identifiers related to one person in a “directory”
• Given one identifier, we can find others
People – Central

- Social Security Number
- Everyone in the world?
- Cumbersome
- Privacy
- Distributed administration

UFID

- Distributed administration of centralized identifier
- 1.8M UFIDs in place
- Over 800 authorized creators
- All creators use one central mechanism for assignment of identifier
- Some duplication. Resolved by hand. 1-2 per day
- UFID used in all enterprise systems, no other disambiguation necessary
Disambiguating people

- Different systems identifier the same person using different identifiers – an employee number and a student number
- Student data and employment data is then combined. Same people appear in both sets
- Disambiguation algorithms identifier duplicates and propose resolutions
- Unnecessary at UF where UFID is assigned first

People – Federated

- Distributed entities produce their own identifiers
- Identifiers are trusted across boundaries
- Examples
  - Passports
  - Driver’s licenses
Identifiers at UF

- UF assigns Mike (the person), UFID (the identifier) 8480-8900
- Mike assigns himself a netid: mconlon
- Netid is associated with UFID, associated with Mike
- Netid forms Kerberos principal: mconlon@ufl.edu
- UF uses the Kerberos principal as an official email address mailto:mconlon@ufl.edu

People – Hybrid

- Some part of the identifier is registered and centrally known
- Some part of the identifier is established locally
- Examples
  - Kerberos principal: mconlon@ufl.edu
  - URLs: plaza.ufl.edu/mconlon
How it Works

• Identifiers are assigned locally
• Institution is known globally
• mconlon@ufl.edu (local/global)
• This identifier is used to identify me across systems
• Systems trust
  – Ufl.edu is a known entity
  – Mconlon is an identifier associated with a particular individual
  – The person who knows the password to mconlon@ufl.edu is the person who was assigned the identifier

Shibboleth – a technology

• Internet2 project with lead site at Ohio State
• NSF, NIH, Microsoft, Elsevier, universities, many more
• Federated identity (multiple identity providers) as well as declarative authorization (attribute release)
• Shibboleth Demo
  http://shibboleth.internet2.edu/demo/shib_demo.html
• See http://shibboleth.internet2.edu
• InCommon Trust Federation
InCommon – a trust federation

- Does every Identity Provider need to trust every Service provider and the reverse?
- [www.incommonfederation.org](http://www.incommonfederation.org)
- Trust federation – each IDP, SP trusts Incommon
- Trust that the federated identity belongs to a particular individual
- 200 institutions, 4 million people
In Action

• UF trusts credentials from other InCommon members
• UF credentials are trusted by other InCommon members
• People can use their local credentials to assert their identity in transactions across boundaries

Works

• Scholarly works – all kinds
  – Books, book chapters
  – Journal articles
  – Presentations (this presentation)
  – Pre-prints, proceedings, posters, abstracts
  – Art, music, photographs, maps
  – Software
  – Cell lines
  – Many more
Identifying works

- The work has a life of its own (it’s a thing), independent from its creators
- The identifiers of the creators are a properties of the work
- “This work is software”
- “This work is created by Mike”
- Multiple identifiers for the same work
- Copies of the work

The citation

- Metadata about the work
- Not copyrightable
- Associates people (authors) with the work
- The citation has an identifier
- Associate citation with person
- Through the citation, associate the person with the work
Person to Identifier to Citation to Work

- Creating an identifier for the work
- Knowing the identifiers of the authors
- Binding a work to its authors in a citation
- Creating an identifier for the citation
- Binding the citation to the scholarly work
- Providing the citation
- Archiving the work
Self-Asserted Association

• Identify the work
• Identify the authors
• Create a citation referencing work and authors
• Trust in the creation of the citation
  – That there is a work
  – That the author identifiers are the identifiers of the people who created the work

What can we do to improve the identification of authors, their works and the association of authors and their works?