Identity and Access Management for the Campus

Benjamin Oshrin
22 July 2019 • Putrajaya, Malaysia
Agenda

- Overview
- Building Identity From Campus Data Sources
- Campus Identity Management
- Translating Campus Identity to the Federation
- Tools and Software
  - Internet2-TIER → InCommon ITAP
Federation services are the tip of the proverbial iceberg for campus identity management. Much infrastructure is needed to support federation.
History of IAM Software in R&E

- DIY
- Commercial Products
  - Not always a good fit, lots of glue
  - Oracle bought Sun and made a lot of institutions unhappy (2009)
- Community Software
    - “InCommon Trusted Access Platform”
Building Identity From Campus Data Sources
Canonical Person Record

- Typically no single, authoritative repository of all people
- Campus Systems of Record ("SORs") authoritative for specific populations
- Data feeds often set up in an ad hoc manner to solve specific problems
  - eg: Issuance of campus ID cards, determining library borrowing privileges, etc
- Campus IAM group usually defines the canonical person record
Canonical Person Record

- IAM group serves as both business and technical interface across authorities
  - Working with various institutional departments and groups to determine specific identity needs
  - Providing the technical services to meet those needs
- Canonical Person Record can also be used to construct the campus directory
- Canonical Person Record is built in the Person Registry, possibly with the assistance of an ID Match Component.
Building the Canonical Person Record

- Step 1: Establish Data Feeds from Campus SORs
- Step 2: Match Records Across SORs, Reconcile Data Discrepancies
- Step 3: Assign Identifiers
- Step 4: Provision Person Record to Downstream Systems/Applications
Where Do Campus Records Come From?

- Campuses typically have distributed authority ("Systems of Record") for person record management
- Separate SORs for different populations
  - Students, Employees, Alumni, Affiliates and Guests
  - Each SOR may have its own policies and procedures for registration
  - Government regulations might apply to some or all populations
    - eg: Work authorization
  - In general, campus SORs are responsible for compliance with campus and governmental policies when registering a new member of the community
## Typical Campus Data Sources / SORs

<table>
<thead>
<tr>
<th>Authority</th>
<th>Domain</th>
<th>Eligibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registrar</td>
<td>Students</td>
<td>Application, Matriculation</td>
</tr>
<tr>
<td>Human Resources</td>
<td>Employees</td>
<td>Employment</td>
</tr>
<tr>
<td>Alumni Relations</td>
<td>Alumni</td>
<td>Graduation</td>
</tr>
<tr>
<td>varies</td>
<td>Guests</td>
<td>Sponsorship</td>
</tr>
</tbody>
</table>
One Person, Multiple Affiliations

- An individual might be multiply affiliated with the institution
  - eg: An employee who also takes classes, or a student with a part time job
- Affiliations can change over time
  - Undergraduate → Graduate
  - Change in job
- Best practice is to create a single person record in the Person Registry for each actual person, regardless of how many or which affiliations
  - Simplifies credential management, auditing, etc
Setting Up Data Feeds (SOR → Person Registry)

- What attributes need to be included?
  - Biodemographic and record matching attributes
    - Name(s)
    - Date of Birth
    - Identifiers (national, SOR, campus, etc)
  - Authorization attributes
    - Department / group information
  - Directory attributes
    - Address and contact information
Setting Up Data Feeds (SOR → Person Registry)

- Frequency of feeds will vary in accordance with SOR capabilities
  - Manual
  - Batch
  - Real Time
- Formats will vary in accordance with SOR capabilities
  - Spreadsheets
  - Flat Files
  - XML
  - APIs
Building the Canonical Person Record

- Step 1: Establish Data Feeds from Campus SORs
- Step 2: Match Records Across SORs, Reconcile Data Discrepancies
- Step 3: Assign Identifiers
- Step 4: Provision Person Record to Downstream Systems/Applications
Record Matching Across SORs

- Record Matching is the process of identifying individuals with multiple affiliations across SORs
- How good the process is depends on the available data
- Conceptually, Record Matching is implemented by the *ID Match Component*, which could be a standalone component or part of the Person Registry
- Record Matching generates a unique *Reference Identifier* per person, which may or may not be used anywhere outside of the ID Match Component and Person Registry
Record Matching Across SORs

- *Can’t we just use the national ID number?*
  - International participants?
  - Lesser requirements for guests?
  - Typos / transcription errors?
  - In some countries, tracking the national ID number is required, in others it is problematic and discouraged
Record Matching Across SORs

- Additional attributes for matching
  - Name
  - Date of Birth
  - Campus Identifiers
    - *Do you already have a NetID?*
  - Email Address (but which one?)
  - Telephone Number (but which one?)
  - Mailing Address (but which one?)
  - Locally available attributes
Reconciling Data Discrepancies

- SORs might have different values for the “same” attribute
  - eg: Preferred vs Official name
  - Errors (eg: date of birth entered incorrectly)
- Errors should be corrected upstream, though this isn’t always possible or easy
- Multiple values should be tracked, with context, policy, or user choice selecting the appropriate value for a given context
  - eg: Self service can show user’s preferred name, while payroll system uses official name
What About Guests?

- *Guest* encompasses a wide variety of use cases
  - Wifi user
  - Library visitor
  - Conference attendee
  - Visiting scholar
  - Parent
Guest Management

• Separate Registration system?
• Part of Person Registry?
• How important is vetting and matching?
  • If a guest returns, how important is it to relink their old identity vs creating a new one?
  • May vary by constituency, service, regulations, etc
  • eg: Campus wifi access for 1 day may have different access policies than a Visiting Scholar present for a semester
• Use of Social/External Identity may facilitate credential management
Campus Identity Management
Building the Canonical Person Record

- Step 1: Establish Data Feeds from Campus SORs
- Step 2: Match Records Across SORs, Reconcile Data Discrepancies
  → Step 3: Assign Identifiers
- Step 4: Provision Person Record to Downstream Systems/Applications
Identifier Management

- Different applications have various requirements for identifiers
  - “Must be alphanumeric, no longer than 8 characters”
  - “Must look like an email address”
  - “Must be a UUID”
  - “Must be an integer”
- The Person Registry must typically maintain a mapping of campus identifiers, each of which will likely have different characteristics
- More discussion of identifiers in eduPerson and voPerson
## Identifier Characteristics

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Persistence</strong></td>
<td>How long is an identifier associated with an entity before it expires?</td>
</tr>
<tr>
<td><strong>Privacy Preserving</strong></td>
<td>Can an identifier be used to correlate records across multiple services?</td>
</tr>
<tr>
<td><strong>Uniqueness</strong></td>
<td>In what context (scope) is the identifier unique?</td>
</tr>
<tr>
<td><strong>Reassignment</strong></td>
<td>Can an identifier be reused (assigned to a new entity) after it expires?</td>
</tr>
<tr>
<td><strong>Human Palatability</strong></td>
<td>Is the identifier human readable?</td>
</tr>
<tr>
<td><strong>Authority</strong></td>
<td>What is the authoritative source of the identifier?</td>
</tr>
</tbody>
</table>
## Common Types of Identifiers (Within Institution)

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Persistent</th>
<th>Privacy</th>
<th>Reassign</th>
<th>Unique</th>
<th>Human</th>
<th>Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platform / Registry (System to System)</td>
<td>Permanent</td>
<td>n/a</td>
<td>No</td>
<td>Within Person Registry</td>
<td>No</td>
<td>IdMS</td>
</tr>
<tr>
<td>SOR ID</td>
<td>Permanent</td>
<td>n/a</td>
<td>No</td>
<td>Within SOR</td>
<td>No</td>
<td>SOR</td>
</tr>
<tr>
<td>General Application (NetID)</td>
<td>Active User</td>
<td>No</td>
<td>Maybe</td>
<td>Within institution</td>
<td>Maybe</td>
<td>IdMS</td>
</tr>
</tbody>
</table>
Group Management

- Groups are commonly used as a factor in authorization systems
  - Relatively simple “in or out” model
  - More complicated models possible with (eg) full RBAC, but less commonly implemented

- Groups can be based on multiple sources
  - “Authoritative” data from SOR
    - Students in Physics 101, Campus Librarians, etc
  - Ad Hoc
    - Lunch time chess discussion club
Group Management

- Group Registry manages group records
  - Can be same as Person Registry or a standalone component
  - Feeds from SORs can populate authoritative groups
  - Authoritative groups can also be the basis of more complicated groups
    - eg: All undergraduates and graduates in the math and physics departments
Credentialing

- Credential = Identifier + Authenticator
- Once a Person Record is created, it is still necessary to tie one or more Authenticators to the record so the user may access services
- Usually done through some form of User Self Service application
- Need some form of initial authentication
  - PIN or code sent to user’s address of record
  - National ID card or other government issued token
  - In person identity proofing
- Re-credentialing (if a user loses their credential) has similar concerns
Credentials

- Most common type of credential is still the password
  - Make it long and drop the character class requirements
- Also: SSH Keys, Certificates, Hardware Tokens, etc
- Social or External Identities?
Lifecycle Management

- A person’s overall status can vary according to their affiliations
  - A student may transition to Grace Period status at the end of a semester rather than simply expiring, to allow access to services over semester break or while transitioning to alumni status
  - An employee might be granted early access to services pending the official start of their employment (eg: a new faculty member, or to work around slow HR processes)
- Expiration and renewal processes are typically implemented by the Person Registry (based on SOR data, or with the User Self Service application)
What About Social/External Identity?

- Social or External Identities can be useful for certain use cases (to avoid issuing local credentials), such as guest access, parental access, etc
  - Maybe even reset of local credentials
- Google, Weibo, etc
- Social/External Identities can be linked to a Canonical Person Record as credentials for access to appropriately designed campus system
  - System must know how to consume these identities and map them to a Canonical Person Record
User Self Service

- Credential management (password reset, etc)
- Set preferred attributes (name, etc)
- Create and manage ad hoc groups
- Sponsor other people
- Delegated Administration
  - Manage campus sub-populations
- Might be implemented by Person Registry or standalone application
Building the Canonical Person Record

- Step 1: Establish Data Feeds from Campus SORs
- Step 2: Match Records Across SORs, Reconcile Data Discrepancies
- Step 3: Assign Identifiers
  → Step 4: Provision Person Record to Downstream Systems/Applications
Provisioning to Downstream Systems

- Provisioning goal is to take data from the Person and Group Registries for use in downstream applications
- Tight coupling: Provisioning direct from Person/Group Registries
  - Fewer components required, but harder to scale
- Loose coupling: Person/Group Registries provision to single target (ESB or Synchronization Engine), which integrates with downstream applications
  - More components required, but more modular
- An intermediate “attribute cache” such as LDAP may be used to simplify application integration
Provisioning to Downstream Systems

- Use of standards like SCIM and SPML would be desirable, but haven’t really caught on

- Additional considerations:
  - Which identifier to provision to which applications
  - Other attributes: official vs preferred name, etc
  - What happens when a person is deprovisioned? Are files/records archived for audit purposes, or when the person comes back?
  - What happens in downstream systems when an already active record is updated (especially merge/split)
  - Where do credential go? (Ideally just the credential store, but sometimes...)
Translating Campus Identity to the Federation
Campus → Federation

- The primary interface between the campus and the federation is the Campus Identity Provider
  - SAML, but maybe OIDC
  - Can be the same as, or built on top of, the Campus SSO
  - Backed by attribute store (usually the directory) and credential store

- Attributes need to map from campus use to federation use
  - eduPerson is the common path for this
  - Even within eduPerson, commonality of attribute usage in real world implementations is not always there
Campus → Federation (Attribute Release)

- Some campuses are better than others when releasing attributes to the federation
- Problematic approach is to treat federated service providers the same as campus application integrations
  - Everything requires special approval, nothing scales
  - Cannot expect researchers from each federated virtual organization to ask for attribute release; VOs will instead migrate to some other identity service
- Better approach is to support REFEDS Entity Categories
  - Especially Research and Scholarship ("R&S")
Assembling a Campus Identity Management System

- “Loosely Coupled Architecture”
  - ie: A bunch of semi-related components
  - For smaller deployments, the number of components can be too much maintenance
  - For larger deployments, separate components increases scalability

- Open Source vs Commercial vs Cloud
  - Currently, all solutions require some “glue”, but OSS probably the least
  - At least one commercial vendor provides support for each OSS component
  - Commercial and Cloud not yet covering 100% of the space reliably
Person Registry

- Historically home grown
- No real commercial options
- Some campuses have tried leveraging COTS solutions, typically end up with a lot of custom code and customizations in front
- ITAP Component: COmanage Registry
  - Under development since 2010
ID Match

- Historically home grown
- Slightly better commercial options, but usually more oriented towards Master Data Management
  - Typically expensive and require customizations
- Can maybe leverage some OSS components, but also more oriented towards Master Data Management
- ITAP Component: COmanage Match
  - Under development since 2018
  - Based on proof of concept implementation in 2013
Guest Management

- Historically home grown, or inelegantly built on top of campus ERPs
  - eg: Campus guests must be registered in the HR system
- Commercial Cloud option via Cirrus Identity (US based) Invitation Service
- ITAP Component: COmanage Registry
Group Registry

- For smaller deployments, can be part of Person Registry or Directory
- For larger deployments, standalone component with more capabilities may make sense
- ITAP Component: Grouper
  - Under development since 2004
Provisioning Engine (Synchronization)

- Seems to be the more common model (vs ESB)
- Many solutions, commercial and open source
  - Many based on old Sun Identity Manager (Waveset), prior to Oracle acquisition
  - ForgeRock was previous US Higher Ed favorite, until Enterprise License costs started to climb
- ITAP Component: midPoint
  - Not developed by Internet2
  - Supports ConnID framework for connector development, allowing connector reuse across Java based synchronization engines
Provisioning Engine (ESB)

- Message based architecture
- Relatively few deployments in the wild
  - While core component is easy to set up, integrations require more work
- No ITAP endorsed component, but plenty of options (OSS, Commercial, Cloud)
  - Apache {ActiveMQ, Camel, ServiceMix, Synapse}, WSO2, Mule, AWS
Credential Store

- Historically Kerberos, more recently LDAP
- Ideally provisioned directly from credential management application (and only to this one store), but practical considerations often get in the way
- MIT Kerberos (KDC), various LDAP options
- No specific ITAP endorsed product
Directory

- LDAP as integration pattern has been around for a while
  - eduPerson, voPerson schemas
- Can be provisioned from Registries, Sync Engine, or ESB
- Directory can also be credential store
- OpenLDAP, Apache DS, 389 (RedHat/SUSE)
  - Long term direction of various LDAP servers may be up in the air?
  - Also (technically) Microsoft AD
  - No specific ITAP-endorsed product
Identity Provider (IdP) and Single Sign On (SSO)

- May or may not be the same component
- Consume campus identity from Credential Store and Directory
- Protocols: SAML (federated), OIDC (up and coming), CAS (enterprise)
- Various OSS and commercial options
  - Shibboleth IdP (Internet2 → Shibboleth Consortium)
  - SimpleSAMLphp (UNINETT, Norway)
  - CAS (Yale → Rutgers → Jasig → Apereo)
  - All three products now have multi-protocol support
- Shibboleth IdP is the ITAP component, but all three projects have traction in R&E
User / Delegated Self Service

- Existing components have some capabilities, usually targeted to their specific capabilities
- Mostly home grown, sometimes based on existing products
- Still much work to do here for a community oriented solution
- Area of interest within ITAP for future work
NRENs/FedOps and the Campus
The Role of NRENs/FedOps for Campus Identity

- Perhaps not a traditional role from NRENs or Federation Operators, but they are well situated
- Provide national guidance for campus identity standards
- National support of community software
  - Sponsor enhancements to existing products to address national requirements
  - Build hosted solutions for campuses
Resources

- COmanage (Registry and Match)
  - https://spaces.at.internet2.edu/display/COmanage/

- eduPerson and voPerson
  - https://wiki.refeds.org/display/STAN/eduPerson
  - https://voperson.org/

- Grouper
  - https://spaces.at.internet2.edu/display/Grouper
Resources

- **Identity Providers**
  - https://www.apereo.org/projects/cas
  - https://www.shibboleth.net/
  - https://simplesamlphp.org

- **InCommon Trusted Access Platform**
  - https://www.incommon.org/trusted-access/

- **ITAP Containers (Docker)**
  - https://spaces.at.internet2.edu/x/fQFbC
Resources

- LDAP Servers
  - https://directory.apache.org/apacheds
  - https://www.openldap.org
  - https://directory.fedoraproject.org
- midPoint
  - https://evolveum.com/midpoint
- MIT Kerberos
  - https://web.mit.edu/kerberos/
- REFEDS Entity Categories
  - https://wiki.refeds.org/display/ENT/Entity-Categories+Home