Columbia University IAM Use Case

Columbia University Overview
The Trustees of Columbia University in the City of New York is a private, nonsectarian, nonprofit institution of higher education whose activities are concentrated at two locations in New York City and extend around the globe. The university provides instruction through 16 undergraduate, graduate, and professional schools. It operates a variety of research institutes and a library system to support its teaching, learning, and research activities. The university performs research, training, and other services under grants and contracts with agencies of the federal government and other sponsoring organizations. The university enrolls approximately 28,221 full-time and part-time students and employs approximately 14,506 full-time employees, including 5,459 full-time faculty members and research staff. Of these, 1,400 hold positions in the arts and sciences, 3,230 hold health science positions, and the remainder hold positions in the other professional schools.

Columbia University HIPAA-Covered Component
Columbia University Medical Center (CUMC), a division of the university located in the Washington Heights section of northern Manhattan, is one of the largest academic medical centers in the United States. It comprises four schools: College of Physicians and Surgeons, Mailman School of Public Health, College of Dental Medicine, and School of Nursing.

CUMC has three primary areas of focus: scientific research, education, and patient care. CUMC offers a wide variety of degrees, certifications, and continuing education in health sciences. Sponsored research, faculty patient care services, tuition, endowment income, patent royalties, and gifts provide the majority of CUMC’s revenues. Approximately 4,102 students are enrolled at CUMC, with a full-time faculty of 2,123, of whom approximately 317 are tenured. Additionally, CUMC’s staff includes 3,571 part-time faculty instructors, 1,152 full-time researcher staff members, 216 part-time researchers, 256 postdoctoral research trainees, and 1,055 postdoctoral clinical trainees. Approximately 65 percent of the full-time faculty and 25 percent of the part-time faculty hold clinical appointments and have admitting privileges at New York-Presbyterian Hospital (NYPH).

Patient care activities include patient visits performed by Columbia full-time faculty through its medical faculty practice plan, as well as clinical, educational, and administration services provided to hospitals and other health care institutions through contractual agreements for services.

CUMC maintains several clinical and education affiliation agreements with other organizations. The most significant affiliation agreements are with NYPH, St. Luke’s-Roosevelt Hospital Center, Bassett Medical Center, and Harlem Hospital. Certain faculty physicians also provide patient care and supervision of residents at NYPH network hospitals and other affiliates. In addition, through interinstitutional "medical service agreements," CUMC faculty provide patient
care in specialty and subspecialty areas at hospitals in the tristate area (New York, New Jersey, and Connecticut) and occasionally in other parts of the country and the world.

In fiscal 2012, the clinical faculty handled approximately 1.8 million outpatient and emergency room visits and participated in instruction and supervision of 660 university medical students and 950 residents and fellows at NYPH. CUMC physicians generated 64,000 NYPH hospital admissions during the year.

Payments for patient-care services provided by the full-time faculty in both institutional and private office settings are derived mainly from third-party payers, including managed care companies (57 percent), Medicare (16 percent), Medicaid (11 percent), commercial insurance (5 percent), and other (1 percent). Direct patient payments make up 10 percent of total payments.

**Relevant Facts to the University and Identity Management**

The university follows a decentralized model, with each of the schools maintaining authority over financial, technical, and administrative decisions. The university maintains a set of policies, which are enforced over the institution. Within the CUMC a separate set of policies govern the HIPAA-protected covered component.

Below are relevant facts:

- As of 2011, the university maintains an endowment of $7.8 billion.
- As of 2012, the university maintains a total asset value of $13.7 billion.
- Over 367,000 identities are managed by the IAM group.
- Over 600 host/service principals.
- Alumni identities are maintained, with permanent e-mail forwarding/provisioning.
- IT is broken up into three main groups; Columbia University Information Technology (CUIT), Columbia University Medical Center Information Technology (CUMC IT), and individual IT groups organized inside the schools and departments.
- There are two Information Security Offices: one within CUIT and one within CUMC IT; the IAM group is located within the CUIT Information Security Office and governs identity management for all of CU, including CUMC.

**IAM Policy and Governance**

*Governance*

The CU Office of Compliance maintains the [Columbia University Compliance Program](#), which comprises three different committees: the Executive Compliance Committee, the University Compliance Committee, and University Compliance Network. None is solely charged with IAM governance; however, any university-wide IAM-related issues would be raised in any of these three groups.

**Executive Compliance Committee (ECC)**

- Directs policy and set the "tone at the top"
• Monitors progress in key compliance areas
• Reviews updates on compliance developments and best practices
• Allocates appropriate resources for compliance areas

University Compliance Committee (UCC)

• Supports the coordination of compliance activities across the university
• Identifies key compliance areas
• Reviews updates on compliance developments and best practices
• Monitors ongoing progress in key compliance areas

University Compliance Network (UCN)

• Serves as subject matter experts on key compliance areas
• Provides information on compliance activities for their area
• Reinforces the concept of the university as a "compliance community"

The Information Technology Leadership Committee (ITLC) consists of the senior IT leaders (directors and above) from all Columbia University schools and departments. ITLC is charged with determining needs, strategies, and collective partnership. The committee participates in policy development, review, and acceptance as well as limited governance functions and helps align IAM business decisions with school/departmental IT technology needs.

CUMC has an official governing committee for specifically managing HIPAA information security and privacy risks. This governing committee is responsible for all HIPAA-specific risks, which would be inclusive of IAM. However, this committee is not charged to specifically handle IAM related policies and issues. The committee is made up of several key executives that are members of the following individuals:

• CUMC Chief Operating Officer
• CUMC IT Chief Information Officer
• CUMC Information Security Officer
• CUMC Controller
• CUMC Chief Financial Officer
• Vice Dean of School of Physicians and Surgeons
• HIPAA Privacy Officer
• General Counsel

Policies
A single policy over the university as a whole governs the use of identity and access management at Columbia. The policy is simple and contains the following elements.

General University Relevant Policies

System Access and Privacy Policy
The sole policy focused on identity and access management for the university-at-large maintains the following:
- User IDs are created only for specified individuals.
- Violations of these practices can lead to suspension of the user ID.
- Administrative user IDs (IDs that are used by the staff and workforce of the university) are active only while the workforce member is affiliated with the university.
- Unauthorized access of another person’s user ID is prohibited.
- Sharing of user ID passwords is prohibited.
- Individual users are responsible for the user ID that was provisioned to them and the security of any passwords used to access the user ID.
- A user ID cannot be used for commercial gain or illegal activity.
- Security administrators will conduct regular reviews of the user IDs, including attempting to easily guess passwords; any weak passwords must be changed immediately.

**HIPAA-Related Policy**

In addition to this single policy, the HIPAA-covered component of the university (Columbia University Medical Center) maintains a single policy related specifically to the security of protected health information (PHI) and maintaining compliance with HIPAA. All HIPAA-related policies are enforced by the CUMC Information Security Office through a number of programs, including an IT Risk Management program focused specifically on multiuser systems, as well as Security Operations.

**Information Access Management and Control**

- Documented procedures must exist for authorizing, establishing, reviewing, and modifying access to electronic PHI (ePHI) systems. This is the responsibility of ePHI system owners and their IT custodians. Access rights must be reviewed on a periodic basis by the system owners.
- Users of multiuser systems will be maintained with unique user identifiers, preferably the same identifier as maintained by the university-wide ID (UNI) or a CUMC-specific ID referenced as a center-wide ID (CWID). Generic logins are permitted when necessary, and their users are documented by system owners and custodians.
- A process for assigning unique user accounts to systems, as well as a process to identify terminated workforce members.
- ePHI access must be limited to workforce members based on need-to-know to complete legitimate work-related tasks.
- All systems with ePHI will be subject to risk analysis and remediation as defined by the System Certification Program. This program is managed by the CUMC IT Risk Management group and ensures that multiuser system risks are identified and managed leveraging the HITRUST framework. This management includes specific components related to access control and identity management.
- All authentication must be encrypted.
- During clinical emergencies a health care professional is permitted to look up ePHI of patients on behalf of another health care provider who is unable to retrieve the information and is caring for the patient.
Identity and Access Management History

The Identity and Access Management team was put together in an informal structure in 1987 as part of the System Administration team. This team was used to automatically provision UNIX accounts within the main university UNIX system, based on bubble cards. In addition, the team provisioned e-mail access to the e-mail services departments, allowing for provisioning of e-mail throughout the university. This group was not called "IdM" but instead "R&D."

In the late 1980s the university implemented Kerberos for authentication services. This Kerberos authentication was the main source of authentication, leveraging LDAP binds, for both authentication and group membership querying. In today's age the Kerberos authentication system is largely front-ended by authentication systems such as WIND, CAS, and Shibboleth.

Over the next decade other groups and many more schools and outside affiliates (such as the Teachers College and Barnard) desired to have access to the shared services already provisioned, mainly e-mail. These affiliates had their own identity sources and originally their own authentication sources. Working with multiple identity sources was a major challenge, and thus a need was born to provide central authentication services and centralized identity management services—services that could be cross-functional between affiliates. An example included Barnard students' need to be able to use the labs at both Barnard and Columbia.

In 2006 the IAM team moved from System Administration to the CUIT Information Security Office due to a major security vulnerability with the student and employee cards and to provide the ability for students to leverage their student cards as meal cards within food court systems. The old card system used the Social Security number as the primary key on the magnetic stripe (magstripe). This permitted any magstripe reader to store the SSN in all systems where the badge readers were installed. The project to replace the card system was run under the newly formed IAM team. The final product was to permit the provisioning of many services from a single onboarding source of printing the card.

The current flow of onboarding cards looks like this:

1. The departmental administrator (the departmental business manager, aka DA) creates a UNI (user ID) using a system called Delegated Identity Administration (DIA).
2. The DA issues a role to the newly created identity. This role, based on the source department and role selected, automatically provisions additional access and updates the card provisioning system within Campus Safety.
3. HR systems are updated with UNI information.
4. E-mail provisioning automatically takes place, although the user has to activate the e-mail address.
5. VPN access is granted.
6. For the proper roles, lab access is granted.
7. For the proper roles, library access, circulation and physical access automatically granted.
8. Later on a DIA and HR reconciliation process kicks off in batch contexts to update the appropriate systems and ensure the identity information is correct.
In addition to the resulting process, a number of additional tools were built to help manage the identity process. These tools included the DIA tool previously mentioned, as well as the following:

- **UNIEDIT**—Provides the ability to adjust many attributes around the UNI, add e-mail forwarding, adjust roles and affiliations, and more. This is the primary UNI administration tool.
- **"WAFFIL"**—A precursor to the current DIA tool. These tools permit authorized sponsors within the university to create UNIs and apply certain roles, which grant access to university services.
- **LDAP**—A white pages directory that provides demographic information, as well as LDAP affiliations used for privilege management throughout the university for specific applications.

The mingling of privileges and demographic information eventually created strategic long-term problems. It was difficult to understand whether an affiliation was granting demographic data, privilege data, or both. Even more challenging, these affiliations were created based on the client's requesting special attributes. Pretty soon hundreds of affiliations were created, building what is referred to as an "affiliation soup"; it was not well understood which affiliations were used for which purpose.

All these affiliations organically changed to include demographic information tied to very specific roles. This relationship created the new affiliations to be leveraged by additional systems throughout the university. For example, an affiliation was created for the Columbia University Medical Center VPN, which was used for both demographic purposes (physical location of the medical center) and privileges (provide access to only medical center employees to the VPN). This project successfully removed 320,000 identities from being able to access the medical center's data network.

### IAM Current State

The IAM team provides the following services to the community, all of which are defined [here](#).

#### UNIX and E-Mail Account Management

Provisioning and deprovisioning of accounts on the primary e-mail (Cyrus) and UNIX (CUNIX) systems happens automatically based on the demographic feed data and affiliations described above.

This service is currently available only for servers maintained by the CUIT UNIX and E-mail Systems group.

#### Implementation Details

- Account information is written in real time to a separate directory server pool, running [OpenLDAP v2](#) on Solaris 9.
- Solaris 9 and earlier servers use locally written PAM (pam_krb54) and nameservice switch (nss_ahpl) modules.
- Linux servers use the standard pam_krb5, pam_access, and nss_ldap modules.
Authentication

Identity management at Columbia revolves around the university network ID (UNI), a unique identifier assigned to each individual in any way associated with the university. The UNI serves as the primary public electronic identifier on campus (even when other identifiers are used internally), and is used for e-mail, printing, computer labs, CLIO (library catalog), PeopleSoft self-service, VPN and dial-up access, and numerous smaller web-based applications.

Enrollment occurs when an individual first becomes part of the Columbia community, usually when the person first arrives on campus, but sometimes earlier. Individuals activate their electronic identity (UNI) by providing personal information.

Implementation Details

A UNI has the following characteristics:

- The form [a-z]+[0-9]+, where the letter components are generally the initials of the individual and the numeric components are sequentially assigned collision numbers.
- It is never reassigned to another individual.
- It may change for an individual under very limited circumstances, such as name change.

Enrollment takes place in real time via WebCreate. The information required varies according to the individual's source feed and may include one or more identifiers (date of birth, SSN, PIN) sent via e-mail. The system is currently provisioned via feeds from HR, the registrar (student information system), and other key departments and receives over 2.4 million authentication service requests per day.

Authorization

The primary source of central authorization information is the demographic feed data described above. An individual may receive different affiliations according to her or his source feed or feeds, including the attributes set for the individual within the feed by the feed administrator. When the individual is no longer a member of the feed, he or she loses the affiliations granted by that feed. This may then result in, for example, loss of access to restricted websites or loss of e-mail service.

A secondary source of central authorization information is group membership on the central UNIX servers (aka CUNIX groups). Users may create and manage their own CUNIX groups, with memberships subsequently reflected in the enterprise directory as affiliations. Expansion and generalization of this service is planned.

Implementation Details

- Central authorization information is recorded as a multivalued attribute affiliation attached to the individual's directory entry.
- Authorization information is not publicly visible and is only made available to approved applications.
- Authorization information can only be obtained via WIND (as described below) or by binding to the directory with a Kerberos IV ticket for the user. The latter method is deprecated and will be replaced when the directory servers are upgraded by late 2006.
Some affiliations contain nested affiliations within them. For example, the VPN access medical center affiliation is actually an aggregate affiliation of over 20 other affiliations that make up what is defined as the Columbia University Medical Center. Users that have any one of these other 20 affiliations will thus implicitly be given the VPN access affiliation.

A large majority of privileges are provisioned this way, which requires the application on the other end to correctly check the affiliation name for its authorization. When a new instructor comes into the Teachers College, for example, that person is automatically assigned the lab user role, VPN access, and instructor affiliations and automatically granted access to those services.

**Enterprise Directory**

**Identification and authorization information via LDAP**

CUIT accepts demographic data from various sources around the university, including the major student (SIS) and personnel (PAC) systems, various university affiliates (Barnard, Teachers College, etc.), and numerous local departments. This information is correlated to generate the UNI for the individual as well as a single entry in the enterprise directory.

This entry contains information such as the individual's name, address, department, and telephone number. In addition, the entry also contains information regarding aspects of the individual related to the source feed, known as **affiliations**. Affiliations are not publicly visible, and may be used for authorization purposes.

Currently, a complicated hierarchy establishes the "best" source for an individual who comes from multiple feeds, and this best source is reflected as the individual's primary role in the directory. However, by late 2006 it is expected that the directory will no longer require a primary role and will simply display all available roles for an individual. All affiliations for the individual are maintained with their entry, regardless of the number of source feeds.

Not all individuals associated with the university may be publicly visible via the directory. Notably, students who request privacy via FERPA as well as alumni with no other connection to the university are only visible to authorized applications. Furthermore, some publicly visible individuals may request the suppression of individual fields within their entry.

**Implementation Details**

- The directory is implemented using [OpenLDAP v1](http://www.openldap.org), which only supports LDAP v2. These servers run on Solaris 9. An upgrade is planned to OpenLDAP v2 (which supports LDAP v3) running on Linux.
- The directory is rebuilt nightly.
- Data from the major systems and affiliates is provided via batch file transfers and is processed nightly. Data from the various departments is provided via the web interface DIA and is processed two or three nights per week.
- Data processing is performed against an Oracle database and a legacy Ingres database using a number of locally developed programs and scripts.
Web-Based Proxy Authentication and Authorization
aka WebSSO/WebISO/Single Sign-On/Reduced Sign-On

The Columbia University WebISO (WIND) allows web-based applications to authenticate anyone with a UNI without the password being passed through the application. According to the needs of the application, an encrypted identifier may be returned instead of the UNI.

For approved applications, authorization information may also be returned in the response.

Implementation Details
- More information about WIND is available here.
- For non-CUIT maintained servers, Apache modules that interface with WIND have been written by CCIT and CCNMTL.
- WIND runs in the Tomcat Java application environment on Solaris 9 servers.

Secure Web Server Authentication and Authorization

Web pages served from the central secure web server (www1.columbia.edu) and other CUIT-maintained servers may take advantage of UNI authentication and affiliation information for authorization purposes using the standard Apache access control mechanism (e.g., via .htaccess).

Implementation Details
- The locally developed open-source Apache module mod_auth_pamacea is used in conjunction with the central UNIX server PAM infrastructure.

Federation

CUIT is participating in several early adoption pilots of federated identity, allowing members of Columbia University to access services maintained by other universities and organizations using their Columbia UNI.

Implementation Details
- Shibboleth

Additional Information
- CUIT Identity Management data flow graph
- CUIT Identity Management servers graph

Challenges

Columbia University is a decentralized institution, as is the IT organization itself, which makes it difficult to centrally administer identities. Our IAM challenges are not limited to the dominant issues discussed below.

There are many systems of record that are leveraged in order to establish full identity, including the student information system, the HR system, DIA, and other sources. Having many sources of identity causes complexities in reconciling identity. In addition, there are many Active Directory domains with no trusts or forests established between them, which creates a
heterogeneous authentication environment where with links between the central authentication services and local authentication sources. This environment replicates authentication domains rather than leveraging a master source, which ultimately inhibits a unified authentication and authorization model for Columbia systems.

Not all systems are tied directly to the central authentication systems (such as CAS or Shibboleth) and instead run their own local authentication, which creates deprovisioning problems when users move between departments or leave the university, as it requires each individual application owner to remove access manually. Two mechanisms help alleviate this problem (but do not fix it):

- **Deactivation Reporting Process.** System administrators can send a list of UNIs to a set of scripts that will kick back a list of UNIs that are "ineligible" based on its affiliation check. This is a batch process.
- **Urgent Termination Process (UTP).** Essentially this is a list serve by which Application Owners subscribe to receive notifications from client managers in HR when terminations take place. This process requires both the HR managers and the system administrators to make manual changes.

A final short list of issues includes:

- A lack of real-time provisioning/deprovisioning (currently done via batch processes)
- A lack of centralized reconciliation (occurs on a system-by-system basis)
- Current directory attributes are confusing and nonuniform
- Inadequate set of user roles to match environmental needs (roles are being "overloaded," which could potentially grant more access than is required)

These are but a few of the current challenges that are being worked out. In addition, a more formal governance posture to help the IAM group make strategic decisions is forthcoming.

**Resources**

For more information on IAM at Columbia University, please consult the following resources.

- Matt Selsky, *Kerberos at Columbia University*, PowerPoint presentation, October 26, 2010
- CUIT IDM architecture
- CUIT IDM history
- The Trustees of Columbia University in the City of New York Consolidated Financial Statements, June 30, 2012 and 2011
- University Compliance
- Columbia University IAM
- Columbia Identity/Access Management, PowerPoint presentation