Frequently Asked Question Series by CREN

Institutional Directories and Repositories

Campuses are increasing their use of computer technology to provide institution-wide services to their communities. As they do so, it is critical to develop and maintain a campus-wide information infrastructure to support these services under varying degrees of security. This FAQ provides information on the use of centralized directory services or repositories to support campus-wide services. This document is intended to educate university staff and administrators including presidents, provosts, and directors of campus information technology.

Published by
the Corporation for Research and Educational Networking (CREN)

Section One:
An Introduction to Directory Services for Decision Makers

1. What is a directory service?

A directory service is a specialized database that contains information about the members of an institution's community and its systems and services. A directory service also may provide security provisions for the use of this information.

A directory service is used to identify the computer systems and online services that a university or college makes available to its members (e.g., faculty, staff, students, alumni). In an impersonal networked environment, a directory service is frequently the only means for determining who the bona fide members of a community are and what computer systems or online services they are allowed to use. The directory information is combined with other credentials such as a password or digital certificate to establish a user's identity (**authentication**) and then provide access (**authorization**) to services.

Directories are designed for access not only by persons, but also by servers and web applications. A directory service can serve as an institution's core source of knowledge about its community of people and systems. If a person or service—such as a printer or Web application server—is not in the directory or an online resource, that person or service, for access and resource purposes, does not exist.

2. How are directory services currently organized on many campuses?

General-purpose institutional databases and repositories are providing many of the functions of a directory service today. While there are important technical distinctions between directories, databases, and repositories, they share the common goal of providing an institutional database that defines the users and systems in an institution's community.

Currently, many institutions rely on a variety of unrelated and unlinked directories to conduct their business. Many of these directories have been implemented within departments and across campuses to support systems and services that cannot be made available without restriction in a larger community. Because so many departments on campus are actively involved in acquiring such systems and services, directory services have become widespread, but localized.

The resulting fragmented approach to authentication and authorization costs the institution money, limits the availability of services, and requires that information about an institution's members be made available on numerous networked databases.

3. Is it possible to have one centralized directory service?

Yes. Directories can be designed to be one logical place for data listings about people or any person or object of interest within the community. Directories can be either one actual physical directory or a virtual meta directory, integrating data from many directories on a daily, twice-daily, or other updating schedule. A directory also may be just an interface to other back-end data repositories.

4. What are the advantages of an institution-wide directory?

With a centrally managed record of its members, computer systems, and services, a university or college can integrate its information technology services more easily. This is true even where no single department is responsible for information technology; that is, at institutions where information technology is being developed and deployed by individuals and departments acting more or less independently of one another.

A centrally integrated directory service will also save money by eliminating the redundant effort required to develop and maintain parallel services. Centrally integrated directory services can also help to ensure that sensitive, personal information about university or college members is reliable and securely maintained, by reducing the number of servers and sites that need to be secured.

Institutional directories enable many applications to use one major directory or meta directory resource for authentication and authorization, rather than having to rely on information coming from many different directories and applications. Institutional directories enable multiple users and applications to access information quickly and simultaneously.

Thus, central institutional directories become pertinent, timesaving applications for two main reasons:

1. They support authenticated secure access, which can support increased levels of existing services.
2. They enable new services to clients while conserving personnel time and resources.

However, it is not trivial to centralize authentication services. A centralized user accounts office is required and users may object to having to deal with a central organization rather than their local departmental administrator.

5. How do I determine whether an institution-wide directory service is right for my institution?

Institution-wide directory services affect all members of a university or college community that use or provide campus computer systems and online services. Planning for the directory service must therefore be a community initiative designed to ensure that members' current and future needs are understood and planned for. An effective planning process will result in community-wide agreement about the high-level goals for a central directory or repository service, including:

- what information about users, computer systems, and online services it contains;
- how that information is secured, maintained, and kept up to date;
- how (with what resources) the directory service is supported;
- what department(s) or individual(s) are responsible for managing the directory service.

Planning will include a technical design component that identifies and prices the hardware, software, and support options that will meet community-agreed goals. Planning may also assess the current provision of directory services and their cost to the institution. This information will be an essential part of any decision to switch over to a more centralized service.

Section Two:
More Information About Directories

6. What is the difference between a directory and a database?

Directories are specialized databases. Directories, more than databases, need to be standards-based, provide high performance for simultaneous access, and be able to be more easily distributed, replicated, and secured.

7. Are there different kinds of directories?

Yes. Some directories perform specific authentication and authorization services, for example, for specific software applications or network operating systems. Some of these directories are those that come bundled with commercial products such as Novell's Directory Service (NDS), Microsoft Active Directory, and Sun/Netscape Directory Server. Other types of directories are NOS-based directories, or Network Operating System Directories, that are bundled with network operating systems.

The directory services described in this document are general purpose and will adapt to general-purpose solutions. They are intended to support institution-wide systems and services.

As these institutional directories evolve, they will likely use a software protocol specific to directories called Lightweight Directory Access Protocol (LDAP). Directories based on this protocol might be used, for example, to store digital certificates.

8. What are the key design features of directories?

Directories

- store information that is frequently accessed; information that is seldom used, such as record transaction data or financial auditing data, should not be stored in a directory.
- access data stored in multiple (distributed) locations as if it were in a single place.
- can be modified; new information fields can be added or deleted as required.
- can be personalized, supporting customized applications and services.

9. What are the core components of a directory service?

A directory service combines hardware and software with institutional policies and procedures and interconnects with many computer networks and information sources. Directory services support the access of resources on a campus, between campuses, and between institutions and services. The core components of a directory service include*:

- hardware, such as network servers;
- server software, such as operating systems, device drivers, and monitoring tools;
- directory software that holds information about community members, computer systems, and online services;
- client software required to access information in the directory;
- policies and procedures for updating, accessing, and securing information in the directory.

(* *Adapted from Howes, T. et al. 1999. p. 4.*)

10. Are there specific software and /or hardware requirements for a directory?

Because directories stand at the center of an institution's information infrastructure, they must be developed based on a variety of standards. The most important standard is LDAPv3, a standard protocol that ensures the directory service can interconnect or communicate with the many different computer systems and online resources across campus. Other hardware and software choices will be based on careful evaluation of an institution's authentication and authorization requirements and its existing IT infrastructure.

11. What data might be stored in an institution's directory?

A central directory service may hold personal information about members of the campus community. A common set of information is likely to be stored for each member (e.g., name, address, phone number, e-mail address). Different sets of more detailed information may then be held for members of specific campus communities. For students, for example, the directory might store information on class load, major field of study, facilities (e.g., lab, computer cluster) that the student is permitted to use. Central directories will also store information about computers and printers (e.g., type, model, processor speed), and online information services. They may also group information, for example, to identify all full professors within the engineering school. Many issues about privacy need to be addressed to ensure compliance with appropriate laws. In the U.S., care should be taken to abide by the Family Educational Rights and Privacy Act (FERPA).

12. How is information in a directory added and updated?

Some information that is stored in a directory can be created and updated automatically by the user. Other information may need to be entered and amended by an appropriate administrator either manually via a Web-based user interface or in batch over the network, for example, from other campus databases. In all cases, the process generally includes authenticating a user to the directory and the verification of appropriate privileges to assure security and integrity of the data in the directory.

13. How is information in a directory accessed?

Directory information is typically accessed by computer systems and online services rather than by individual users. For example, when a student swipes an ID card through the card reader at the door of a secure facility, the card reader will likely access a directory (or database) to verify that the student should gain access. Similarly, when a faculty member attempts to read an online journal article that is only accessible to members of specific departments, the service supplying the journal article online will access the directory to determine whether the particular faculty member should be allowed to see the article. In the case of use by online services, a directory entry combined with the use of a password by the user is generally part of the authentication process.

As more and more campus activities become dependent upon the authorization and authentication functions of the central directory, institutions will want to take steps to ensure they are always up and running. Some, for example, will keep several copies of the directory running in case one or two go down or fail somehow to handle an abnormal surge in demand.

14. Does an institution need an institution-wide directory to begin using digital certificates?

No. Most applications now in development make use of digital certificates for simpler and more basic identity authentication and authorization needs, such as signed (non-encrypted) e-mail. All that an institution needs to issue digital certificates is a database that *defines the users in the community* so that the service used to issue certificates can determine whether certificate requesters are in fact eligible for such certificates. Incidentally, this database does not need to be published—accessible to the world—nor does it need to use any particular technology, such as the LDAP directory protocol.

An institutional directory is also not required for certificate users to identify themselves to a Web site. However, the Web site may need access to a database that authorizes the user to access the Web site's information. Again, this database does not have to be public, nor does it require any specific technology. Indeed, many campuses currently maintain such databases within their registrar or personnel offices. If these databases are updated and synchronized frequently (e.g., daily), they can be used for issuing certificates that will identify users to a Web site.

Finally, a directory is also not a requirement in order for users to exchange signed secure e-mail messages, such as S/MIME messages. Most (and perhaps all) S/MIME products include copies of users' certificates in the messages they sign.

15. Are there any applications that require directories?

The primary application for which directories are required is the exchange of *encrypted* S/MIME messages. For this exchange, the sender needs to obtain a copy of the recipient's digital certificate that contains the public key to encrypt the message—prior to sending the message. Most products that send S/MIME messages are designed to be able to query an LDAP directory. Additional infrastructure that can support the sharing of a single certificate across multiple machines as well as key recovery is most likely also required for the support of encrypted S/MIME. While it is possible to use encrypted S/MIME in the absence of a directory, it is cumbersome to do so.

Section Three:
Designing and Implementing Directories

16. What about designing directories? Where is more information available?

Many excellent resources about directories and related issues, such as certificate authorities, policies and profiles are all linked at the HEPKI site at **www.educause.edu/hepki**. HEPKI stands for the Higher Education Public Key Infrastructure and is a group linking middleware initiatives from Internet2, Net@EDU, and CREN. Two related sites are of particular interest. The first is the site at **www.educause.edu/eduperson/**.

This site contains the work of the EDUCAUSE/Internet2 eduPerson task force that is defining an LDAP object class called "eduPerson" that includes widely-used person attributes in higher education. A second site is the "Recipe for Configuring and Operating LDAP Directories" at **www.georgetown.edu/giia/internet2/ldap-recipe/**. This LDAP recipe site is an excellent starting point for detailed technical information about schema design for higher education directories.

Section Four:
Glossary

X.500
X.500 is an information model and suite of protocols for defining directory services. One of the protocols in the suite is Directory Access Protocol (DAP), and it specifies how a requester can perform operations against a data store and where that data store is.

LDAP
LDAP stands for Lightweight Directory Access Protocol. It was designed in 1992 because the larger protocol, DAP, made it difficult to create a complete DAP implementation, even client-side implementation, on personal computers at that time. The LDAP data and namespace model is similar to X.500. The primary difference is that LDAP is a simplification of DAP that runs directly over the TCP/IP network stack.

Namespace
A namespace is how information is referenced in a directory. A properly designed namespace allows for easy updating of information in the directory, and flexibility in setting access control and replication policies.

Registry
A registry is a tool that supports the needs of applications for coordinated maintenance of identity information about a class of business objects. Some people use the word "meta directory" interchangeably with registry. The registry can be thought of as the identity reconciliation function.

For example, we may want to know if Susie Q is the same as Susan Quinlan, since Susie Q is a student and Susan Quinlan is an employee. A critical function of the directory is the identity reconciliation function. It is knowing that Susie Q is or is not Susan Quinlan, and knowing that in the student system she is known by an ID, in PeopleSoft she is known as Student/Administration-ID-such-and-such, and in the Human Resources department she is known as Oracle-Financials-person-such-and-such. The registry supports mapping across those systems of interest within the directory.

Within a registry, directories can derive value-added information by synthesizing and coalescing information that comes from disparate source systems. The registry is responsible for passing along changed data, such as an updated telephone number, to the (perhaps) multiple copies of the directory that exist so that all copies of the directory are synchronized. All individuals related to the campus should be included in the registry.

Schema
Schema are guidelines that determine what data can be stored in a directory. Schema designs help to maintain the integrity and quality of data and provide a well-documented, predictable way for directory-enabled applications to access and modify the collection of directory objects.

There are several sources for standard schema formats, including LDAP, X.500, and industry consortium standards. A project called eduPerson is working on a schema for higher education directories.

S/MIME
S/MIME, which stands for Secure/Multipurpose Internet Mail Extensions, is a set of official Internet specifications for securely sending different types of mail objects, such as text, image, audio, or video, over the internet. The latest version of the specification is S/MIME v3 as of November 2000, and it is now proceeding through the standards track of the Internet Engineering Task Force (IETF).

Section Five:
References

Gauthier, Larry. "Directory-Enabled Computing: The Directory's Expanding Role." Midvale, Utah: The Burton Group, 1999. 47 pages. www.tbg.com.

Grewe, Frank and Michael LaHaye. "Directories on Campus: Getting Started." CREN Webcast on November 4, 1999. Available at www.cren.net/know/techtalk/events/getstarted.html.

Hodges, Jeff and Frank Grewe. "Campus Directories." CREN Webcast on April 22, 1999. Available at www.cren.net/know/techtalk/trans/directories_1.html.

Howes, Timothy; Mark C. Smith and Gordon S. Good. *Understanding and Deploying LDAP Directory Services*. USA: MacMillan Technical Publishing, 1999.

Klingenstein, Ken and Keith Hazelton. "Building Directories: The Fundamentals." CREN Webcast on February 17, 2000. Available at www.cren.net/know/techtalk/trans/directauthen_1.html.

Schneier, Bruce. *Secrets and Lies: Digital Security in a Networked World*. New York, NY: John Wiley and Sons, Inc, 2000.

Authors, Reviewers, and Editors

George Covert
Associate Director
Academic Information Technologies
Iowa State University

Michael Gettes
Lead Application Systems Integrator
Georgetown Institute for Information Assurance
Georgetown University

Frank Grewe
Manager
Enterprise Internet Services
University of Minnesota

Keith Hazelton
IT Architect
University of Wisconsin

Jeff Schiller
Network Manager of Information Systems
Massachusetts Institute of Technology

The initial source of many of the questions in this FAQ is Howard Strauss of Princeton, the Technology Anchor for the CREN TechTalk series. Also many thanks to Brian Leney of the Council on Library and Information Resources for his assistance in the production of this FAQ.

CREN is a nonprofit, member-based organization dedicated to supporting professionals in higher education and research organizations with strategic information technology knowledge services and communications tools. CREN's services include seminars, workshops, and a series of FAQs on strategic technology areas combined with practical implementation strategies. CREN offers certificate authority services supporting secure access to online resources and interinstitutional resource sharing. CREN also produces a series of TechTalk webcasts that feature higher education guest experts in an informal format, discussing current and emerging technology trends and topics.

CREN

1150 18th Street NW, Suite 1030, Washington, DC 20036
Phone: (202) 293-6838 ● E-mail: cren@cren.net ● Web: www.cren.net