Planning Directory Services

A Findlay & D Mahl

Euroview Project

The EuroView Project was supported by the European Commission under the Fourth Framework. The partners were Dr Materna GmbH (DE), Brunel University (UK), and Sema Group sae (ES).

1. Introduction

Directories are becoming an essential component of IT strategy. However, the territory will be new for most managers and few staff are likely to have the appropriate skills. This document aims to assist planners by discussing the important issues and providing guidelines on the deployment process from start to finish.

The issues associated with directories span a broad range and relate to the legal, technical and organizational. The skills called for during the deployment effort will be diverse and are unlikely to be provided by any one group. Consequently, the process may well be collaborative and contributions will certainly be made by various parts of the organization. Legal issues and database management aspects, for example, should be the preserve of administrators, while technology issues should be in the domain of the IT department. As such, involvement of all relevant groups must be foremost in the planners mind. It is certainly necessary to exercise a degree of sensitivity as the service will have a great effect on the way staff work and the way the organization operates as a whole.

1. This Document

This document forms the first in a series relating to the deployment of a corporate directory service. Its aim is to provide a basis for the development of a directory strategy. Beginning with requirements analysis, the design, implementation and maintenance phases are subsequently discussed.

The second document in the series is targeted at implementers and will contain a more detailed account of the process.

2. About Directories

A directory is a way of finding information associated with a name. In the paper phone book, for example, people find telephone numbers by looking up the relevant person's name.

Electronic directories work in the same way. The directory consists of a set of structured names that map to entries in the database. Like the paper directory these names can be used to retrieve their associated information. Where the electronic version differs is that names can be automatically searched for and that many different types of information can be stored. This means that the database is rich in content and that data is easily accessible.

A number of different types of directory probably already exist in your organization. The internal phone book is an obvious case. Others are the human resources database and possibly electronic address books used by your e-mail system. It is important to note that many different types of organizational data are already available, and are accessible in different ways.

Electronic directories act to unify organizational data. They provide a focus and a single access point for the information. In addition to making directory data available to local users, electronic directories can be distributed so that the information of multiple organizations can be shared. The long term aim of directory strategists is to see a single, unified, global, directory.

As well as accessing directory data users can also be allowed to make updates to the database over the network. Because updates made in this way modify the source database the scope for data integrity is much greater. This contrasts with paper, where corrections can only be made available each time the directory is published. In addition this means data management responsibilities can be assigned in a more meaningful way. System managers can maintain e-mail

information, switchboard staff can keep telephone data up to date and users can maintain their own personal information. This flexibility is the key to the directory's ability to provide consistently useful service in a rapidly changing information framework.

3. Benefits To Your Organization

Without being able to find out the address of a person, you can't reach them. This initial step in the communication process is poorly served by existing network services. The directory aims to fill this gap by providing an efficient ''one stop'' access point for all forms of contact information. If you want to find someone's e-mail address you can find it in the directory. If you need find the phone number of a service, it's there.

Because the directory is searchable contact points can be located by their real world name. Name the person you want, and the addressing information will be returned to you directly. Because the source database is accessed directly, you'll always have the latest, up to date information at your fingertips.

This benefits an organization by promoting efficient internal communication, and, importantly, the way in which citizens and clients contact and communicate with the enterprise. Your services are easier to reach because contact points can be discovered in a structured, user friendly and convenient manner.

The gains are more than just practical. Publication of useful information is good for public relations, because it signifies that your organization is open to communication with the citizen, business and government. In this way the directory gives your organization a network presence and accessibility that spans international borders. The importance of this cannot be overstated given the ever increasing scale and diversity of the global networked community.

4. Planning A Directory Service

The directory can be a focal point for information services if implemented correctly. The success or failure of a deployment project will be influenced by a range of factors. As such it is better to step back and get a wide, balanced viewpoint as early as possible.

Begin by looking at what you want the service to do. The directory was designed to support services that require addressing information - anything from fax and e-mail to video

conferencing and public key management for security systems. Your specific requirements will stem directly from the applications you currently provide and aim to provide in the future. The directory will provide supporting infrastructure for these. How accessible the directory is and how well it integrates with these services will be the key and deciding factor.

Defining exactly what you want to achieve is essential, as this will clearly delimit what you'll have to do. However, in-built flexibility is always useful, as determining the requirements for a service as potentially wide ranging as the directory will be difficult!

Once you've decided on the desired functionality you can move on to the practicalities. Some of the more important considerations are:

- Your legal responsibilities.
- Selecting the appropriate technology.
- Security policy.
- Procuring the right software.
- Gathering the data and converting it to populate the directory.
- Managing the data and maintaining the service.

5. Determine The Requirements

Who is it for?

At an early stage decide whether or not the directory will be available for external access. If external use is deemed a requirement, the next step is to determine whether any information intended for staff use is sensitive and should be excluded from public access. It may be appropriate to define a public view containing the list of official contacts, and a private view encompassing the entire database.

At this point security policy will also have to come into consideration. You should aim to keep policy as simple as possible because implementing and maintaining a complex set of rules will result in significant extra overhead. The assignment of access rights to internal (private) and external (public) users on a per-entry basis is relatively straightforward. The overhead will be much greater if access controls are to be applied on a per-attribute basis. An example of this would be to

hide the telephone of certain people within the organization, whilst allowing their e-mail addresses to be visible.

Complex security policies can be based on the notion of select directory communities. This might be a set of organizations that co-operate closely with you. However, once you enter into this territory you'll encounter policy problems such as how communities are defined (the question of who is trusted and should be within a community is not always simple) and technical problems, such as the need for secure methods of authentication between community members.

Access controls are generally applied within the directory server, i.e. access information is held in the directory database. This approach requires the use of encrypted authentication techniques in order to be truly meaningful. However, recent developments allow the flow of directory data to be limited using firewall software. These directory proxy servers act as an access filter and prevent sensitive data leaving your local network.

What will it do?

Consider how people will use the directory to find the information they want. Will users simply want to find contact information by looking up personal names? If the service is to be valuable to the citizen then contacts must also be available by keying on service names and descriptions. If the services provided have international scope then language translations of service descriptions could be useful.

Go on to think of specific functions that you want the directory to perform. In general the basic role of the directory is to provide a White Pages service by publishing the contact information of organizational staff. A directory of services is the next step. Other functions are possible. A simple expansion of the directory database utilises the directory entries of people to manage electronic distribution lists. Usefulness is not limited to the support of network services. The directory information model is extensible, and so the database can be used to store data of local interest. Examples are user identifiers, payroll information and organizational data such as job titles.

The flexibility of the directory is an advantage but can also be a pitfall, as there may be a temptation to over specify the requirements. Functionality that is difficult to implement might hinder progress. Be careful to identify functions that are practically achievable within limits of time and resource. Those that aren't can be left to a later date.

How will Users Access the Service?

The way users access the service will contribute to the success of the project. If appropriate and usable tools for accessing the directory aren't provided, the service won't be used! It is thus important to find clients that are user friendly, practical and efficient.

Key access methods can be identified early on by looking at the ways in which people access information services now. The World Wide Web is an obvious candidate for service delivery, and this can be provided through use of a directory to Web gateway. The advantages a Web gateway has is that little or no software roll out is required, as most desktop computers will already have a browser installed. In addition gateway access means that the directory service can be integrated into the organization's Web site, and so be utilized as a means of service provision to clients and citizens.

Though Web access is useful, the telling factor for local staff will be how well the directory is integrated with existing applications. Integration with the e-mail interface is of primary importance. This is closely followed by word processors and other office applications. More advanced use of the directory can be made by group calendar and address book applications.

System integration is the next step. The data that the directory provides can be used as the basis for a number of applications. The directory is best thought of as an information resource that can be used to support existing and future applications. As an example, the database can be used to manage a document hierarchy or a World Wide Web page index. White Pages information can be used to manage further communications services, such as video conferencing. For video conferencing, personal entries can be used to point location services to where a user is logged in, and hence reachable from.

Thinking about the applications that the directory could support in the future is a good idea. Anticipating future changes will result in a plan aimed towards a more open and flexible design.

6. First Steps

Identify the Data Sources

By now you'll have a clear idea of what information you want to put in the directory. The initial move toward achieving this is

locating where that data is currently held in your organization. The telephone and e-mail databases are almost the certainly the first stop. You might also want to consider the human resources database.

Now would be the right time to consider some management issues. Firstly, how will the directory be placed in the organizational infrastructure? Will it become the master source for some of the databases already mentioned? It might make sense to master telephone numbers in the directory. The e-mail database will almost certainly be dependent on the mail system in use. The directory may not provide the level of database functionality required by a human resources department (the directory, for example, cannot replace a relational database). The introduction of the directory can be an opportunity to rationalise existing databases and their associated management functions.

Whatever you decide, bear in mind that your plans will require the co-operation of people currently in charge of the databases. Doing what you can to involve them in the project early on could make life easier for you later!

Start Thinking About Data Management

The next priority is to find suitable ways of merging the information in each of the data sources. The information content and format of these databases is likely to vary considerably and so this will present a problem. Tools for converting the source databases into the format required by the directory may have to be written specifically.

Once conversion tools have been created, tools for merging the data streams will need to be found. This issue is both difficult and highly important. Reliable merging will be required for the ongoing management task, but is likely to problematic due to inconsistencies in the database. Much of the contention will be due to variations in naming - in one database a person may be named ''J Smith'' and in another ''John Smith''. Similarly, how can the merging tools tell the difference between two different people with the same name and in the same department? The question is far from trivial and much investment may have to be made in order to maintain an adequate level of consistency in large databases.

In addition to the ''bulk'' loading of data, procedures for small changes will have to be defined. These procedures will take care of individual changes, for example when new staff arrive, leave or change posts. In cases where the directory is not the master source of data, synchronisation tools that automatically

update the directory in response to changes in the master data should be considered. Synchronisation tools may also be needed to directory data is used to populate external databases, such as proprietary address books.

Assess your probable management need by making estimates of how big the database is likely to be and how much change due to staff movement and re-organization will occur each month.

Your Legal Responsibilities

Though data protection and privacy laws vary across European nation states, most of the basic tenets are the same. These are founded in the rights of the individual (which is the primary concern). Specific data that should be excluded is that related to racial or ethnic origin, political opinions, religious beliefs, trade-union membership, and data concerning health or sexual life. Common sense dictates that personally sensitive data serves no purpose in the directory should not be published. Further laws relate to people's right of access to their information, specifically over the integrity of data held and the right of refusal.

The practical effects on the manager of a directory service are that the mechanisms for supporting the following user requests must be included:

- To view all information stored about themselves.
- To demand restoration of erroneous data.
- To refuse being included in the directory.

7. Design The Service

Design the Database Structure

User considerations should be paramount. The database should reflect a natural organizational structure and be partitioned in a logical way. Service contact points could, for example, be held in one database subtree in order to make browsing easier. Entries in the database should be named in a readable and descriptive way, rather than using impenetrable acronyms or identifiers.

It may be advantageous to keep public and private views distinct in the database, again by placing the data in separate access areas. This makes sense from the perspectives of security

and management as it eliminates inter-dependency. Implementation of the security mechanism in particular will be simplified by ensuring that public data is kept apart from private data.

Service Infrastructure

The infrastructure design will determine how effectively the service is rendered. The design should thus reflect the level of service required. If your organization is small then the service will be handled adequately by a single directory server. In certain situations more than one server will be required, an organization with several sites lacking strong network interconnections being the most obvious example. Here low-bandwidth connections would limit the usability of the directory if one machine at a central site was used to service all requests. In this case, the solution is to install servers at each of the sites and implement a data replication scheme so that each server holds a copy of the entire directory database. In this way every local user has fast access to a site-local access point, with no need to resort to inter-site networking when resolving directory requests.

8. Procurement

Server Software

Make sure that interoperability and adherence to standards is high on the list of priorities when assessing software solutions. If interoperability is of sufficient quality then you should be able to select from a range of products. This means, for example, that you'll have the freedom to purchase server and client software from different vendors. Similarly you'll be able to choose products that match available platforms, which will be useful if your organization is a mixed hardware environment. Most vendors submit their software to interoperability tests and you should ask for the results of these. Software developers should also have a commitment to future standards and advances in technology. It won't help if the current product is conformant with prevailing standards, but future offerings diverge from the straight and narrow.

Ask for demonstrations that show how well the product performs in stressful conditions, e.g. with large databases. The ability to run a search quickly through a large number of entries is important, as is the speed at which directory updates

are processed. Request performance figures from tests using databases at least as big as your own.

Client Software

Interoperability is again important. The freedom to mix and match is fundamental. As well as allowing you to select different products to match different needs, it means that other organizations can access your directory using their software. Interoperability improves accessibility.

Desktop integration is vitally important. User interfaces will ideally work with the applications staff are using now, e.g. word processors, calendar/diary managers and e-mail interfaces. Clients that employ non-proprietary methods of integration are advisable as these will co-operate more easily with standard software.

It may be appropriate to look at clients specific to a application. E-mail interfaces in particular may have an associated directory component that is capable of accessing your service. Latest versions of the popular communication suites from Netscape, Lotus and Microsoft are examples.

Equipment Needs

Begin by estimating levels of usage, using this to determine the hardware requirement. It's probably better to overestimate somewhat in order to allow for future growth.

Other factors are the level of resilience required and the overall size of the database. If resilience is an issue then backup machines may be needed.

It may be possible to combine a directory service with other services on the same machine, but the demands of each service must be considered if service levels are to be maintained.

9. Service Roll Out

Awareness and Training

Ensuring that users are fully appreciative of what the directory can do is important, as this will help the service achieve it potential and maximize return. Awareness campaigns and training materials will be needed to accomplish this. Much of this will simply be published material, though some face to face training of help desk staff will be necessary. Friendly and

intuitive user interfaces will help of course, though a degree of familiarity through experience is often needed before functionality can be properly explored.

Installing Software

Client software will have to be installed on user machines. If a Web gateway is in use then timescales can be extended, as simple access will be catered for. Integrated desktop applications with higher functionality will require some installation effort.

10. Maintaining The Service

The directory is like many other information services in that the long term maintenance costs will probably outweigh the initial effort. Much can be done to reduce maintenance overhead by examining the issue at an early stage. Formalising the requirement early on will be helpful. How up to date does the directory have to be? How tolerant to short term errors are you? If tolerance is low then procedures for management of the database should be defined well before the directory is in operation.

Consider the purchase of dedicated tools for synchronisation of the directory with source databases. A good synchronisation utility will be able to detect changes in a source database and make those changes to the directory. The aim here is to avoid regular complete downloads of the master database into the directory, as this could have a negative affect on service performance.

A management interface for manual update of the data will also be useful. Such a tool should exist for administrators to make small changes to the directory, e.g. a change of telephone number when a person moves office.

11. Cost Analysis

Hardware and Software Costs

The number of machines and software licenses required will depend on the topology of the directory infrastructure and how much backup is required. Assuming the requirements dictate response times of a few seconds and twenty four by seven availability then backup for every server containing master data will be necessary.

A single directory server should be able to cope with a database consisting of several thousand entries. Though this, of course, is influenced by the power of the hardware and the particular implementation used. As a very rough guideline expect to pay about 6000 ECU for a hardware platform (including OS costs and other support software) that will support 10,000 entries. A server license will cost between 750 and a 1500 ECU, although the cost is likely to scale downwards as license numbers increase. For a single site, one main and one backup server would put the server cost at around 10,000 ECU. Scaling up linearly, a larger organization with five sites may need to spend

Be careful to include client costs in the capital expenditure. Client software should be relatively inexpensive and may well be as little as 10 ECU per head under a site licensing scheme. Again, costs will vary and depend on numbers.

As well as user access software, you'll need to consider the requirement for update interfaces that will be used by administrators to manage the directory database.

Ongoing Management

Database integrity will demand a consistent level of manpower. One or more people dedicated to the task of managing the service may be necessary. The exact level of effort needed will depend on the size and nature of the organization. Estimating this will be difficult and so it maybe useful to postpone the formation of a management team until some experience has been gathered. Be careful though, as this approach might lead to a overloading of existing staff in the initial phases, while data is verified through usage. In all cases procedures for directory update should be defined. This means advertising a contact point used to channel requests for change and assigning responsibility for the task. The management duty is extremely important and should not be underestimated!

Data management will involve software costs. Data synchronisation tools, if available, will require expenditure. If suitable tools are not available, then a degree of in-house effort to develop, for example, data conversion tools could be called for.

Getting further information in 2002

This document was first published in 1997. Other documents in the EuroView series cover government directory service requirements in more detail. Most are based around the idea of 'White Pages' service, though since the end of the project in 1998 the requirement has moved more towards using directories to drive networked services. Products such as Novell's NDS, Microsoft's Active Directory, and IBM's Secureway all use directory technology, but they emphasise the database and system management benefits of directories rather than the name-and-address lookup ones.

It is quite possible (and may well be desirable) to serve both requirements from a single directory, thus contributing to a simplification of information management.

In the past few years, many of the smaller players in the directories market have merged or disappeared, though some new ones have sprung up. Some very strong Open Source projects have made excellent progress, supplying directory servers and clients at no cost per installed copy. Similar projects have extended the work into the areas of security and Public Key Infrastructures.

Most of the staff involved in the EuroView project have now moved on, but the deliverables are still available from the Brunel University website:

`http://www.brunel.ac.uk/x500/euroview/`

Andrew Findlay is now an independent consultant specialising in large-scale systems management, networks, security, and directory services. He maintains a website at:

`http://www.skills-1st.co.uk/`

and can be contacted as follows:

Dr A J Findlay
Skills 1st Ltd
2 Cedar Chase
Taplow
Maidenhead
SL6 0EU

+44 1628 782565

andrew.findlay@skills-1st.co.uk