

# Request For Comment: Proposed Public Health Directory (phDir) Projected Capacity Design

**Stephen H. Fishman**

## Revision History

Revision 1	9/11/2002
------------	-----------

## Introduction

### Table of Contents

[Purpose](#)  
[Participation](#)  
[Communication](#)

## Purpose

The purpose of this document is to solicit input regarding the estimates and specifications contained here in related to the planned Public Health Directory (phDir) for use by the public health community.

The phDir is the directory services initiative at the Centers for Disease Control and Prevention (CDC). The phDir will provide critical contact information and other data about Public Health professionals to enhance communication within the Public Health community. When designing a directory service it is important to identify all of the factors that will ultimately impact how the directory will perform in production. This document will cover specifications around three factors that have a large performance impact on a directory. First, the ultimate size of the directory to be supported, second the amount of requests to be supported, and lastly the performance benchmark goals that to be supported. It is only after these specifications are in place that informed architecture decisions can be made.

A draft capacity plan and network architecture will be formed from the information gathered as a result of this request. This capacity plan is intended to serve two purposes:

1. To be the implementation capacity plan for the Public Health Directory (phDir) which is being implemented at CDC.
2. To act as a possible reference model for LDAP directories in public health.

The proposed plan contains many assumptions regarding the size and daily activities of the public health population. Comments regarding these assumptions from the public health community are vital in properly designing a system that will fulfill the needs of the public health professional.

Please note that while there are many planned uses for the public health directory, described in other documents, the primary use for the initial release of the directory is to serve as a central "white pages" directory for common lookup activities. It is with this use in mind that this capacity plan is being composed. As new needs for the directory service are identified (e.g., an automated alerting system), at any time within the life of the directory service project or after the directory service has gone into production, it will be important to revisit this both this capacity plan and other physical system architecture documents to determine if the proposed or actual architecture can support the additional request load that would be incurred by that new need.

## Participation

Recipients of this document are invited to submit their comments, questions and recommendations concerning this capacity plan. The initial comment period ends on 10/21/2003. After that date, a version 1.0 capacity plan will be released with a network architecture. Evolution of the capacity plan and associated network architecture is expected based on knowledge of future uses for the directory server such as authentication and authorization for CDC applications. Input received after the comment period ends will be considered in the development of the subsequent version of the capacity plan.

## Communication

Comments may be submitted via email to either <[sfishman@cdc.gov](mailto:sfishman@cdc.gov)> or <[rtanzola@cdc.gov](mailto:rtanzola@cdc.gov)>. All responses should clearly indicate the section of this document to which they pertain. After the responses have been collected, a new draft of the document will be produced including a revision log to indicated changes for each version of the document.

# Chapter 1. Directory Sizing Specification

### Table of Contents

[Overview](#)  
[Public Health Community Estimates](#)

## Overview

The core of the phDir directory will be comprised of entries that represent public health professionals, organizations to which they belong, and roles that they perform. This section starts the process of estimating both the size of the public health community and the behaviors that the public health community engages in on a daily basis. These initial estimates provide the basis by which emergency, growth and request-type scenarios can be formulated.

# Public Health Community Estimates

The estimates presented below were done with input from several members of the public health community. Several definitions are included here in an effort to clarify the nature of the data presented in this document.

## *Organization Type*

This column includes a comprehensive categorized listing of every type organization that would have a use for and people recorded within the directory.

## *Number of Organizations*

This column estimates the number of organizations that populate the organization category.

## *Directory Users per Org*

This column estimates the average number of people recorded within the directory per organization. This number should include public health personnel as well as people necessary to support the public health mission indirectly (e.g. - technology workers supporting a HAN, etc.)

## *Directory Users per Org Type*

This column is a calculation equal to Number of Organizations x Directory Users per Org

## *Directory Requests per Directory User per Day*

This column estimates the average number of requests a person recorded within the directory for this type of organization will make over the course of a day.

## *Requests per Day per Org Type*

This column is a calculation equal to Directory Users per Org Type x Directory Requests per Directory User per Day

The size and behaviors of the public health community are estimated below.

**Table 1.1. Public Health Community Estimates**

<b>Organization Type</b>	<b>Number of Organizations</b>	<b>Directory Users per Org</b>	<b>Directory Users per Org Type</b>	<b>Directory Requests per Directory User per Day</b>	<b>Requests per Day per Org Type</b>
State Health Departments	50	250	12,500	10	125,000

Organization Type	Number of Organizations	Directory Users per Org	Directory Users per Org Type	Directory Requests per Directory User per Day	Requests per Day per Org Type
County Health Departments	3,000	25	75,000	7	525,000
City Health Departments	6	50	300	7	2,100
Universities	120	10	1,200	3	3,600
Small Federal Gov't Agencies	20	50	1,000	8	8,000
Large Federal Government Agencies	7	3,000	21,000	12	252,000
Hospitals	10,000	15	150,000	2	300,000
Clinics	50,000	5	250,000	1	250,000
International Organizations	200	20	4,000	5	20,000
Territories	9	25	225	5	1,125
NGO's	50	20	1,000	3	3,000
Professional Organizations	30	20	600	7	4,200
<b>Totals</b>	<b>63,492</b>	<b>N/A</b>	<b>516,825</b>	<b>N/A</b>	<b>1,494,025.00</b>

Some other estimates that will affect the number of entries in the directory are as follows:

1. on average each organization will have 10 subunits requiring 634,920 additional organization entries.
2. there will be 5000 standard roles supported by the organization

## Chapter 2. Request Load Specification

### Table of Contents

[Overview](#)  
[Request-Type Specifications](#)  
[Future Projections](#)

## Overview

The targeted audience of the phDir directory is comprised of public health professionals, administrators, emergency personnel and people who support the activities engaged in by the preceding groups. This section provides more detailed estimates of the size, growth projections and behaviors engaged in on both a daily and emergency basis by the public health community. These estimates provide the

basis by which architectural decisions can be made in relation to how to support the needs of the public health community. The request scenario estimates are presented below. In preparing these estimates, several assumptions are being made:

- a serious public health emergency could escalate the request volume by a factor of 50
- any widely distributed client application will not be tree oriented in its interface
- the directory service will not be open to querying outside of approved client applications

## Request-Type Specifications

Further details on public health community behaviors are presented below.

**Table 2.1. Request Scenario Estimates**

Scenario	Requests per Day	Requests per Hour	Requests per Minute	Requests per Second
Normal Operations	1,494,025.00	62,251.04	1,037.52	17.29
Emergency Operations	74,701,250.00	3,112,552.08	51,875.87	864.60

Along with the size of the public health community and the volume of requests being made it is also important to identify the type of requests being made to the directory. Write operations, and membership queries require significantly more processing power than simple requests for a particular person or organization known by a distinguished name (DN). Assumptions are being made here as to what percentage of requests from the public health community fall into four different categories: Simple Query, Complex Query, Simple Update and Complex Update. Further definition of each of these categories is provided below.

The four different categories from the previous section (Simple Query, Complex Query, Simple Update and Complex Update) are defined from a user perspective. Each one of these request types has different impacts upon the directory service and need to be taken into consideration when making design decisions regarding the directory service schema.

The first category, Simple Query, is defined as any query that from a user perspective is intended to return a particular user within the directory. A possible scenario might be one where a user needs to get contact information about another user, enters several characters from the first and last names and the home state of the user being searched for. For the purposes of this document, a Simple Query will be assumed to make on average 1 request to the Directory Service.

The second category, Complex Query, is defined as any query that from a user perspective is intended to return either a group of users within the directory who are members of an organization or conversely is intended to return all of the groups to which a particular user is a member. A possible scenario might be where a user needs to get send out a communication to all the members of one or more organizations. Technically, this would involve making 1 query for information about the particular group and then making subsequent queries to retrieve member entries. For the

purposes of this document, a Complex Query will be assumed to make on average 10 requests to the Directory Service.

The third category, Simple Update, is defined as any update that from a user perspective is intended to update the information stored for one particular user within the directory. This may however include multiple points of data to update. A possible scenario might be one where a user needs to update their contact information to reflect their new wireless phone as well as to modify the entry that manages the priority of delivery methods. For the purposes of this document, a Simple Update will be assumed to make on average 3 requests to the Directory Service.

The fourth category, Complex Update, is defined as any update that from a user perspective is intended to update the information stored for many users within the directory as a result of some organizational change. A possible scenario might be one where a user needs to update the membership information for all of a particular group's members due to a name change within an organization or a re-organization of personnel. For the purposes of this document, a Complex Update will be assumed to make on average 25 requests to the Directory Service.

**Table 2.2. Request-Type Estimates - Normal Operations**

Request-Type	% Of Total	User Requests per Day	User Requests per Hour	User Requests per Minute	User Requests per Second	LDAP Requests per User Request	LDAP Requests per Second
Simple Query	75 %	1,120,518.75	46,688.28	778.14	12.97	1	12.97
Complex Query	15 %	224,103.75	9,337.66	155.63	2.59	10	25.94
Simple Update	6 %	89,641.50	3,735.06	62.25	1.04	3	3.11
Complex Update	4 %	59,761.00	2,490.04	41.50	0.69	25	17.29

Again, the assumption being made here is that a serious public health emergency could escalate the request volume by a factor of 50.

**Table 2.3. Request-Type Estimates - Emergency Operations**

Request-Type	% Of Total	User Requests per Day	User Requests per Hour	User Requests per Minute	User Requests per Second	LDAP Requests per User Request	LDAP Requests per Second
Simple Query	75 %	56,025,937.50	2,334,414.06	38,906.90	648.45	1	12.97
Complex Query	15 %	11,205,187.50	466,882.81	7,781.38	129.69	10	1,296.90
Simple Update	6 %	4,482,075.00	186,753.13	3,112.55	51.88	3	155.63
Complex Update	4 %	2,988,050.00	124,502.08	2,075.03	34.58	25	864.60

Request- Type	% Of Total	User Requests per Day	User Requests per Hour	User Request s per Minute	User Request s per Second	LDAP Request s per User Request	LDAP Request s per Second
Update*							
<b>* - It is unlikely that the frequency of complex updates will increase as a result of a public emergency</b>							

## Future Projections

Growth in traffic can be projected in two dimensions. First, growth in the public health community. Second, growth in the applications that utilize the directory. We will first present calculations estimating the growth in the public health community. The assumption being made here is a 12.5% growth factor in public health community staff.

**Table 2.4. Five Year Growth Estimates In Requests per Second**

Operations Status	Requests per Second 1 Year Out	Requests per Second 2 Years Out	Requests per Second 3 Years Out	Requests per Second 4 Years Out	Requests per Second 5 Years Out
Normal Operations	19.47	21.91	24.65	27.73	31.19
Emergency Operations	973.72	1,095.43	1,232.36	1,386.41	1,559.71

Once again, using the same percentages presented previously this data can be broken out according to the type of request.

**Table 2.5. Five Year Growth Estimates In User Requests per Second by Request-Type**

Request- Type	% Of Total	User Requests per Second 1 Year Out	User Requests per Second 2 Years Out	User Requests per Second 3 Years Out	User Requests per Second 4 Years Out	User Requests per Second 5 Years Out
Simple Query	75 %	14.61	16.43	18.49	20.80	23.40
Complex Query	15 %	2.92	3.29	3.70	4.16	4.68
Simple Update	6 %	1.17	1.31	1.48	1.66	1.87
Complex Update	4 %	0.78	0.88	.99	1.11	1.25

**Table 2.6. Five Year Growth Estimates In Requests per Second by Request-Type in Emergency Status**

Request-Type	% Of Total	Requests per Second 1 Year Out	Requests per Second 2 Years Out	Requests per Second 3 Years Out	Requests per Second 4 Years Out	Requests per Second 5 Years Out
Simple Query	75 %	730.29	821.58	924.27	1,039.81	1,169.78
Complex Query	15 %	146.06	164.32	184.85	207.96	233.96
Simple Update	6 %	58.42	65.73	73.94	83.18	93.58
Complex Update*	4 %	38.95	43.82	49.29	55.46	62.39

**\* - It is unlikely that the frequency of complex updates will increase as a result of a public emergency**

**Table 2.7. Five Year Growth Estimates In LDAP Requests per Second by Operations Status**

Operations Status	Total LDAP Requests per Second 1 Year Out	Total LDAP Requests per Second 2 Years Out	Total LDAP Requests per Second 3 Years Out	Total LDAP Requests per Second 4 Years Out	Total LDAP Requests per Second 5 Years Out
Normal Operations	66.80	75.15	84.54	95.11	107.00
Emergency Operations	3,339.86	3,757.34	4,227.01	4,755.38	5,349.81

**Figure 2.1. Five Year Growth Estimates In LDAP Requests per Second by Operations Status**



To estimate request growth over time for the next five years we must also consider the growth in applications that use the directory for authentication and authorization. Assumptions are being made here both for the number of users accessing CDC applications and as to the number of application requests which fall into three different categories: Simple Authentication, Simple Authorization, and Complex Authorization. This document assumes 13,000 individual people on average will use CDC applications every day and that that number will grow 12.5% every year. Further definition of the three categories is provided below.

The three different categories from the previous section (Simple Authentication, Simple Authorization, and Complex Authorization) are defined from a user perspective. Each one of these request types has different impacts upon the directory service and need to be taken into consideration when making design decisions regarding the directory service schema.

The first category, Simple Authentication, is defined as any attempt to bind to the Directory Service from a single user.

The second category, Simple Authorization, is defined as any attempt to determine if a user has access rights to a specific application.

The third category, Complex Authorization, is defined as any attempt to determine specific data, screen or field level rights within an application.

The assumption being made for the purposes of this document is that over the course of a year the number of applications that require authentication and authorization services to be supplied by a directory will grow by 5.

**Table 2.8. Authentication and Authorization Requests 1 Year Out**

Request-Type	Number of Users	Number Of Applications	Request per Application	Request per User per Day	Request per User per Second	Request per User per Second - Emergency Status
Simple Authentication	13,000	5	4	20	3.01	150.46
Simple Authorization	13,000	5	20	100	15.05	752.31
Complex Authorization	13,000	5	100	500	75.23	3,761.57

**Table 2.9. Authentication and Authorization Requests 2 Years Out**

Request-Type	Number of Users	Number Of Applications	Request per Application	Request per User per Day	Request per User per Second	Request per User per Second - Emergency Status
Simple Authentication	14,625	10	4	40	6.77	338.54
Simple Authorization	14,625	10	20	200	33.85	1,692.71
Complex Authorization	14,625	10	100	1,000	169.27	8,463.54

**Table 2.10. Authentication and Authorization Requests 3 Years Out**

Request-Type	Number of Users	Number Of Applications	Request per Application	Request per User per Day	Request per User per Second	Request per User per Second - Emergency Status
Simple Authentication	16,453	15	4	60	11.43	571.29
Simple Authorization	16,453	15	20	300	57.13	2,856.45
Complex Authorization	16,453	15	100	1,500	285.64	14,282.23

**Table 2.11. Authentication and Authorization Requests 4 Years Out**

Request-Type	Number of Users	Number Of Applications	Request per Application	Request per User per Day	Request per User per Second	Request per User per Second - Emergency Status
Simple Authentication	18510	20	4	80	17.14	856.93
Simple Authorization	18510	20	20	400	85.69	4,284.67
Complex Authorization	18510	20	100	2,000	428.47	21,423.34

**Table 2.12. Authentication and Authorization Requests 5 Years Out**

Request-Type	Number of Users	Number Of Applications	Request per Application	Request per User per Day	Request per User per Second	Request per User per Second - Emergency Status
Simple Authentication	20,283	25	4	100	24.10	1,205.06
Simple Authorization	20,283	25	20	500	120.51	6,025.31
Complex Authorization	20,283	25	100	2,500	602.53	30,126.57

## Chapter 3. Benchmark Specification

### Table of Contents

[Overview](#)

## Overview

When designing a directory service it is important not only to determine the ultimate load in both size of directory you want to support and amount of requests to be supported but also to specify the performance benchmarks of such a system. It is only after both of these specifications are made that informed architecture decisions can be made. This section provides benchmarks for what is an acceptable amount of time for the Directory Server to fulfill a request.

## Request Benchmark Specifications

Minimum performance standards for the planned Directory Service are presented below.

**Table 3.1. Request Benchmarks**

<b>Request-Type</b>	<b>Acceptable Response Time</b>
Simple Query	.25 sec.
Complex Query	1 sec.
Simple Update	.50 sec.
Complex Update	2 sec.
Simple Authentication	.25 sec.
Complex Authentication	1 sec.
Simple Authorization	.50 sec.
Complex Authorization	2 sec.