



# ARChER 1.0 Overview

<b>ARChER 1.0 Overview</b> .....	<b>1</b>
About the ARChER Toolset .....	2
About this document .....	2
Overview .....	2
Development Partners .....	2
ARChER Toolset functionality .....	2
Which components do I need? .....	4
Testing .....	4
Downloads .....	5
Future Development of the ARChER Toolset .....	5
Technical details .....	6
Overview .....	6
Detailed architecture diagram .....	6
ARChER Data Services .....	8
Overview .....	8
Configuration .....	9
Pre-requisites .....	10
Configuration .....	11
Network requirements .....	13
Overview .....	13
ARChER Servers .....	13
Inbound access .....	13
Outbound access .....	14
Detailed network diagram .....	15
Additional access during installation .....	15
Distribution server .....	16
Installing the ARChER Toolset .....	18
Overview of steps .....	18

## About the ARCHER Toolset

---

### About this document

This overview is a guide to the ARCHER Toolset 1.0. The first section provides non-technical information. The second section is more technical and includes detail intended for system administrators and network managers.

### Overview

The ARCHER Project<sup>1</sup> completed with delivery of the ARCHER Toolset, a set of e-research enabling software. ARCHER was an Australian higher education initiative which has developed “production-ready” software tools, operating in a secure environment, to assist researchers to:

- collect, capture and retain large data sets from a range of different sources including scientific instruments
- deposit data files and data sets to e-research data repositories
- populate these e-research data repositories with associated metadata
- permit data set annotation and discussion in a collaborative environment, and
- support next-generation methods for research publication, dissemination and access.

The ARCHER Toolset is a suite of third-party and ARCHER-developed components which assist the collection, storage, management and publication of data from scientific instruments.

The released version of the ARCHER Toolset is V1.0, released as Open Source GPL3. No further development will be conducted within the ARCHER Project.

Development of most of the ARCHER Toolset components continues outside ARCHER within the open source community. Links are provided on this website at <http://archer.edu.au/products>

### Development Partners

The ARCHER project partners were Monash University<sup>2</sup> (lead institution) in Melbourne, The University of Queensland<sup>3</sup> in Brisbane and James Cook University<sup>4</sup> in Townsville.

### ARCHER Toolset functionality

The four applications of the ARCHER Toolset are:

Component	Purpose
DIMSIM & CIMA	Distributed Integrated Multi-Sensor & Instrument Middleware, a plugin for the Common Instrument Middleware Architecture, allows concurrent data capture and analysis from instruments.
XDMS	Enables exploration of the storage facility, deposition of data by project, addition of metadata to data and datasets and export to external research repositories.
HERMES	Enables client desktop management of the transport of large data files and datasets to

---

<sup>1</sup> <http://www.archer.edu.au>

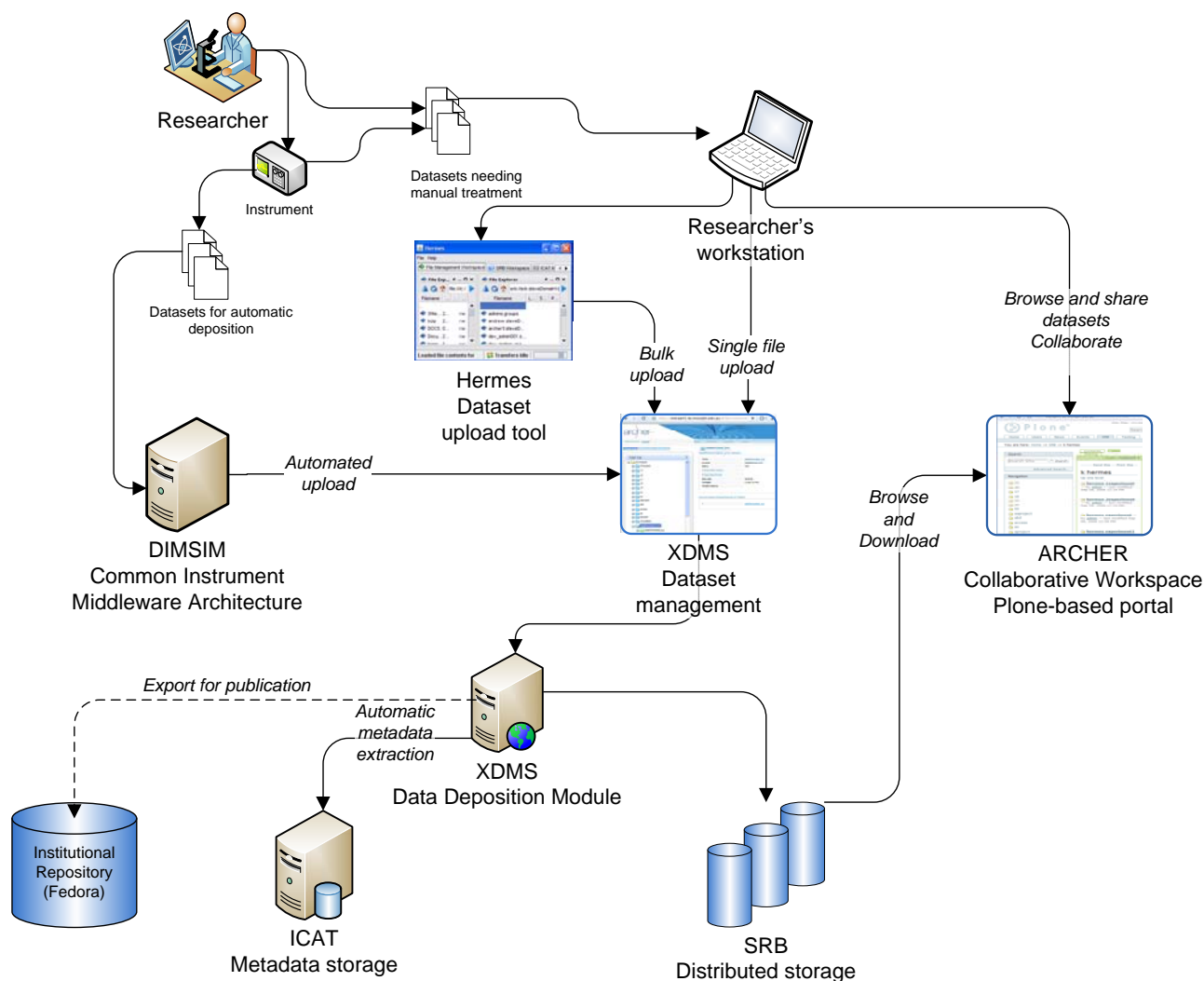
<sup>2</sup> <http://www.monash.edu.au>

<sup>3</sup> <http://www.itee.uq.edu.au/~ereseach/projects/archer/index.php>

<sup>4</sup> <https://ereseach.jcu.edu.au/projects/archer>

	an e-research repository.
ARCHER Collaborative Workspace	Provides an authenticated portal and document repository for discussion, comment and pre-publication collaboration between researchers, including access to the repository of e-research data

The components fit together as follows:

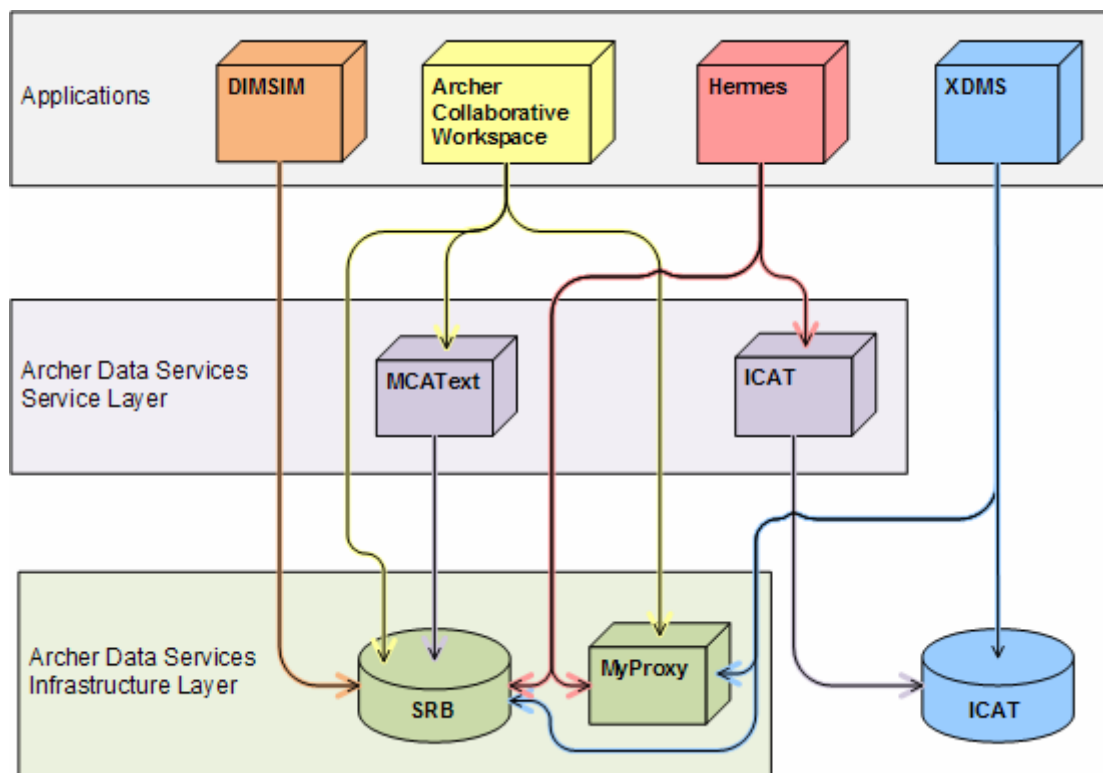


These four applications are supported by two layers of software infrastructure: the ARCHER Data Services.

Layer	Purpose
ARCHER Data Services Service Layer	Provides web service interfaces to the data and metadata storage.
ARCHER Data Services Infrastructure Layer	Provides distributed storage, authentication, and an optional certificate authority.

## Which components do I need?

The modular architecture of the toolset means you do not have to install the entire toolset. The following diagram shows the main dependencies:



**High-level architecture diagram**

**Note:** SRB and the ICAT database have collectively been known as the “ARCHER Research Repository” in some publications.

Each of the four user applications are independent. However:

- ARCHER Data Services Infrastructure Layer provides the SRB storage and is required for every ARCHER application.
- ARCHER Data Services Service Layer is required by ARCHER Collaborative Workspace.

**Note:** The ICAT database (not to be confused with the ICAT service) is installed as part of XDMS.

The system requirements for each component are given in the corresponding system administrator’s guide.

## Testing

The initial version of the ARCHER Toolset (ARCHER V 1.0) has been subjected to installation, operational and functional testing. The results of this testing are available from the ARCHER website. Critical and major defects found during testing have been predominantly addressed within the constraints of time and funding. However, not all of the related fixes have been incorporated into the ARCHER 1.0 release.

The known defects are presented as a table from the ARCHER Testing at <http://archer.edu.au/products>.

## **Downloads**

All components of the ARCHER Toolset can be downloaded from

<http://www.archer.edu.au/downloads/>.

## **Future Development of the ARCHER Toolset**

The ARCHER Toolset V1.0 is released as open source specifically to enable the customisation of the components to suit varying research areas.

As the ARCHER Project has completed, further development of the toolset components devolves to the wider community wishing to facilitate eResearch.

Those developing any component further should note the copyright and contact details included in the source code of the initial release.

At October 2008, continuing development of ARCHER Toolset components is known to include resolution of defects of importance found previously but not yet resolved.

Those undertaking additional development are invited to collaborate in the spirit of open source by providing later versions to the wider community which include updated documentation.

## Technical details

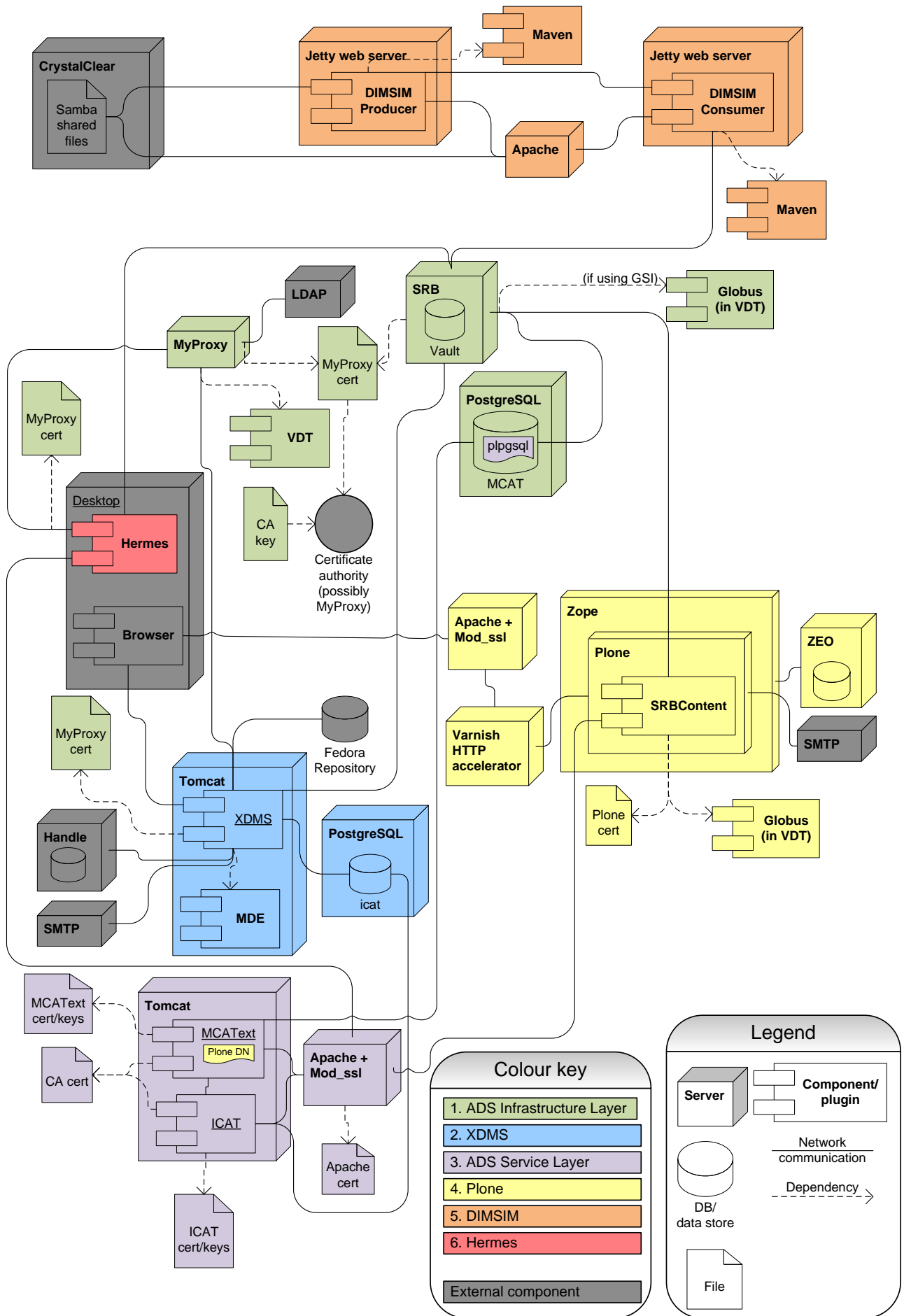
---

### Overview

This section is intended for system administrators who will be installing ARCHER Data Services and the ARCHER application suite. It gives an overview of the infrastructure and service components, how they fit together, and how to install them. To install and maintain these tools, familiarity with Linux, grid technologies, X509 authentication, PostgreSQL, Apache web server, and Apache Tomcat is required.

### Detailed architecture diagram

The following diagram shows the relationships between the main components in the ARCHER Toolset. Components connected by a solid line must be able to communicate via network. Components connected by a dashed line must be located on the same physical server.



Detailed architecture diagram

Components such as Tomcat and Apache do not need to be installed before installing the ARCHER Toolset. The individual installation instructions cover their installation, if they are not already present.

## ARCHER Data Services

### Overview

The ARCHER Data Services (ADS) infrastructure supports other components in the ARCHER Toolset, and incorporates both third-party and ARCHER-developed software. It provides data storage via SRB, metadata storage with ICAT, and authentication through MyProxy.


ADS is made up of two layers:

<b>ARCHER Data Services Infrastructure Layer</b>	
SRB	<p>SRB, or Storage Resource Broker, is a layer of infrastructure that provides a secure grid file storage system. It is a commercial product, free for use by academic institutions, and with full source code available. One or more network file systems normally provide the data storage.</p> <p>More information:</p> <ul style="list-style-type: none"> <li>• <a href="http://www.vislab.uq.edu.au/research/SRB/background.html">http://www.vislab.uq.edu.au/research/SRB/background.html</a></li> </ul>
VDT	<p>VDT is a grid software packaging system that installs several grid components, including MyProxy and the Globus libraries.</p> <p>More information:</p> <ul style="list-style-type: none"> <li>• <a href="http://vdt.cs.wisc.edu/">http://vdt.cs.wisc.edu/</a></li> </ul>
MyProxy	<p>MyProxy is a service that listens on port 7512, generating short-lived certificates called "proxies" for users upon request. It serves these over the network, allowing users access to remote services. ARCHER adds LDAP authentication to the standard MyProxy installation.</p> <p>More information:</p> <ul style="list-style-type: none"> <li>• <a href="http://grid.ncsa.uiuc.edu/myproxy/doc.html">http://grid.ncsa.uiuc.edu/myproxy/doc.html</a></li> </ul>
CA	<p>Many parts of ARCHER Toolset require a common Certificate Authority (CA) to sign certificates. ARCHER Data Services can set up MyProxy to function as a CA. This is useful in a testing or development environment.</p> <p>In a production environment, this is normally a government or university-run service, which matches a person or institution's credentials with a certificate. Setting up or gaining access to a certificate authority in a production environment is beyond the scope of this document.</p>



ARCHER Data Services Service Layer	
ICAT	ICAT is a metadata storage service that implements the CCLRC Scientific Metadata Model Version 2 <sup>5</sup> to record information about scientific experiments. The data from the experiments itself is stored on the SRB, while the metadata is held in the ICAT. The ICAT's storage is implemented as a PostgreSQL database, which is installed through the ARCHER XDMS application. The ICAT webservice layer is part of ARCHER Data Services.
MCAText	MCAText is an ARCHER-developed web service layer over SRB and its MCAT database. It provides a high performance mechanism for other services to lookup authorisation information on content within SRB. It provides update notification to other services when content is modified, moved, or created. It is used by certain ARCHER tools, including the ICAT service and ARCHER Collaborative Workspace.

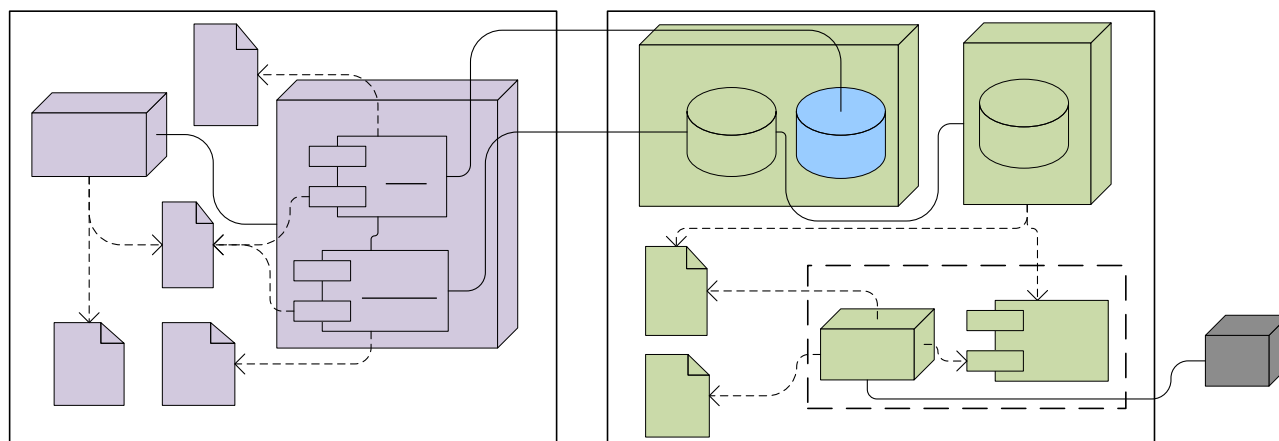
These two layers provide the basis for the ARCHER applications DIMSIM, XDMS, ARCHER Collaborative Workspace, and Hermes.

Note	
	<p>The third party components that make up ADS Infrastructure Layer are installed with ARCHER-specific customisations. This means you cannot install ARCHER applications with off-the-shelf installations of components such as SRB or MyProxy.</p>

## Configuration

Several different configurations of the various components (CA, VDT, SRB, MyProxy, ICAT, MCAText...) are possible. The ARCHER documentation focuses on the configuration that was tested by the ARCHER team. Two machines were used:

- The "back" server hosts the ADS Infrastructure Layer: SRB, ICAT database, MyProxy (as CA), and connects to an external LDAP.
- The "front" server hosts the ADS Service Layer: ICAT and MCAText web service, which are accessed through an Apache server to provide security.



**ARCHER Data Services tested configuration**

<sup>5</sup> <http://epubs.cclrc.ac.uk/bitstream/485/>

Other configurations are discussed in the ARCHER Data Services Infrastructure Layer System Administrator's Guide.

### Pre-requisites

The ARCHER Data Services installation assumes freshly built servers. All testing was performed on this kind of environment.

Installed following the layout given in this document, ARCHER Data Services requires the following:

<b>Back server (Infrastructure Layer)</b>	
Operating system:	Centos/Red Hat Enterprise Linux/Scientific Linux 5.2 or later.
Free disk space:	Applications: 3 Gb SRB storage: As much as possible.
Memory:	1 Gb RAM.
External services:	Access to an LDAP server
Installed software:	Subversion. ("yum install subversion").

<b>Front server (Service Layer)</b>	
Operating system:	Centos/Red Hat Enterprise Linux/Scientific Linux 5.2 or later.
Free disk space:	2 Gb free disk space.
Memory:	1Gb RAM.

Requirements for ARCHER applications such as XDMS and DIMSIM are given in their respective System Administrator Guides.

**Note:** This document assumes that the operating system is Red Hat Enterprise Linux, or Centos. All ARCHER and third-party components other than SRB are theoretically platform-independent, and could be installed on other Linux distributions or even Microsoft Windows, but this has not been tested and is beyond the scope of this document.

## Configuration

The standard, tested configuration services on each server as follows.

<b>ARCHER component</b>	<b>Front server</b>	<b>Back server</b>
DIMSIM	Jetty, Apache	Jetty
ADS-IL – MyProxy		MyProxy (7512)
ADS-IL – SRB		SRB (5544) MCAT (PostgreSQL)
ADS-SL – ICAT	Apache + mod_ssl Tomcat	(ICAT, installed by XDMS)
ADS-SL – MCAText	Apache + mod_ssl Tomcat	(MCAT, installed by SRB)
XDMS	Tomcat	ICAT (Postgresql)
ACW	Apache	Varnish Plone Zope ZEO



## Network requirements

---

### Overview

This section is a summary of the network requirements of the standard, tested ARCHER configuration. It is intended for use by a network manager to configure firewall settings and to assist in maintaining the services that are running.

### ARCHER Servers

Two servers are assumed in this configuration.

- The “front server” primarily hosts Tomcat and Apache.
- The “back server” primarily hosts databases and authentication services.



**Both** servers must be accessible to end users.

Other configurations using more servers are possible. See the ARCHER Data Services System Administrator's Guide for more information.

### Inbound access

The following table lists all the services that are installed and the ports they listen on. It assumes that all components are installed with default settings according to the installation instructions.

#### Front server:

Server	Address	Port	ARCHER component
Apache with SSL	https://server/mcatext https://server/icat	443 443	ADS-SL
Apache	http://server/plone http://server/zope	80 80	ACW (Plone) ACW (Zope)
Jetty	http://server:8081	8081	DIMSIM (consumer)
Tomcat	ajp://server:8009/mcatext ajp://server:8009/icat	8009 8009	ADS-SL ADS-SL
Tomcat	http://server:8080/xdms	8080	XDMS

#### Back server:

Server	Address	Port	ARCHER component
PostgreSQL	jdbc:postgresql://server/icat	5432	XDMS, ADS-SL
	jdbc:postgresql://server/MCAT	5432	ADS-IL, ADS-SL
Zope/Plone	http://server:8080/<plone site name>	8082	ACW
Varnish	http://server:8081	8081	ACW
Jetty	http://server:8080	8080	DIMSIM (producer)
Apache	http://server	80	DIMSIM (producer)
SRB	srb://server	5544	ADS-IL
MyProxy		7512	ADS-IL

**Note:**

On the back server, MyProxy and SRB need to be accessible by users. All other services are only used by applications running on the front server.

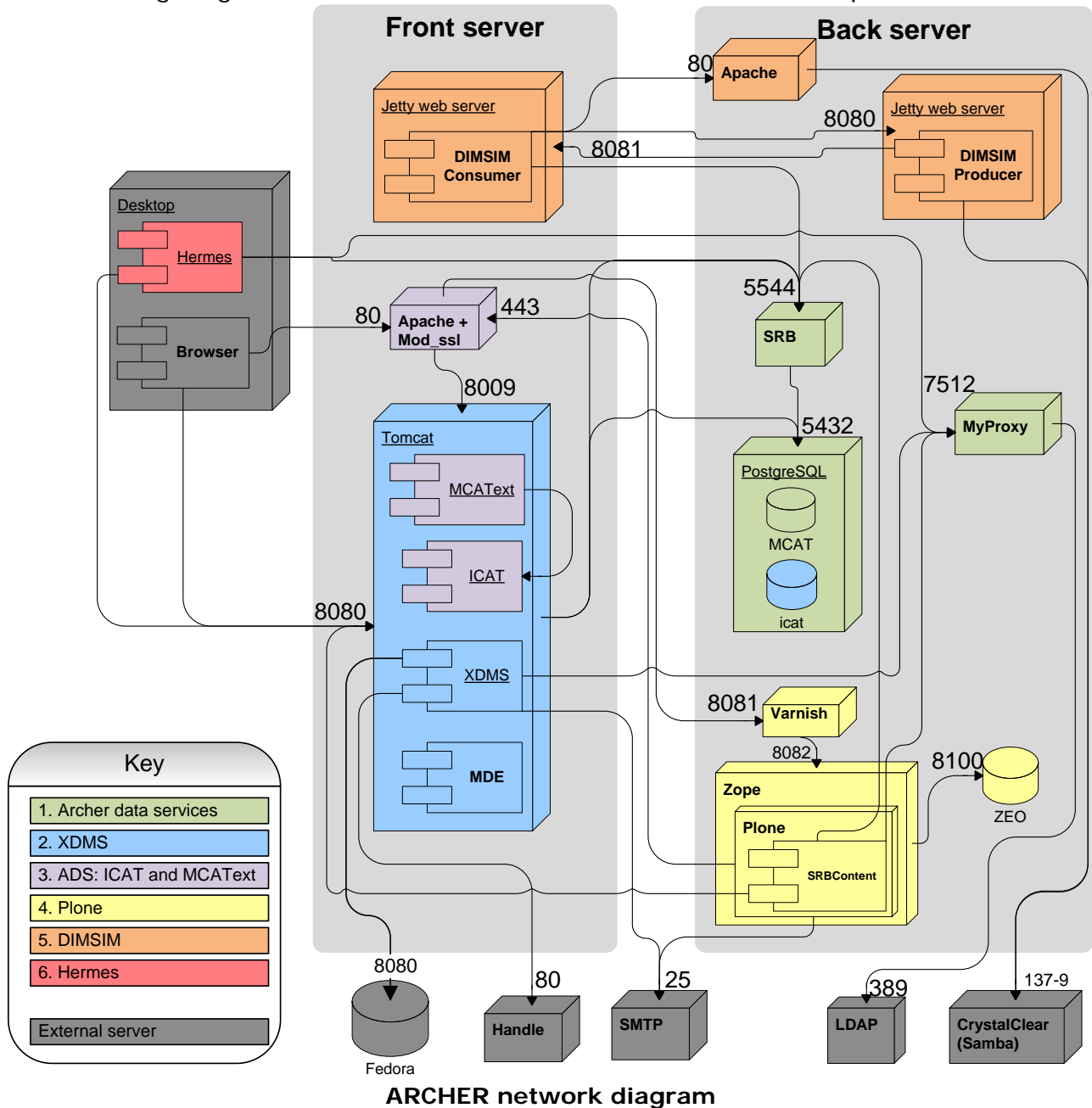
**Outbound access**

Access to certain external servers is required by individual ARCHER components.

<b>Server</b>	<b>Component</b>	<b>Purpose</b>	<b>Typical port number</b>
LDAP	ADS IL	(Optional) Provides authentication for MyProxy.	389
SMTP	XDMS, ACW	Sends mail.	25
Fedora	XDMS	(Optional) Stores exported datasets in a repository.	8080
Samba	DIMSIM (Producer)	Contains CrystalClear datafiles to be ingested.	137-139
Handle	XDMS	(Optional) Assigns persistent identifiers to datasets.	80
Shibboleth	XDMS, Hermes	(Optional) Provides single sign-on.	80

## Detailed network diagram

The following diagram shows all connections between ARCHER components.



## Additional access during installation

### Outbound

During installation, outbound access is required as follows. This does not include access required to download the installation packages themselves.

Component	Access required
ADS-IL	<ul style="list-style-type: none"> <li>• Yum repository</li> <li>• "Distribution host" (JCU or other)</li> <li>• VDT cache.</li> </ul> <p>See "ARCHER Data Services Infrastructure Layer System Administrator's Guide" for more details.</p>
ADS-SL	None required.

XDMS	None
DIMSIM	May require access to external Maven repository.
ACW	<ul style="list-style-type: none"> <li>• Yum repository</li> <li>• Access to web, to download Zope, Plone and third-party Plone products.</li> <li>• JCU Subversion repository</li> </ul>

## Inbound

During installation, inbound network access is required for ARCHER Data Services Infrastructure Layer. The back server must have a DNS name that can be resolved externally. That is:

- "host <fully qualified server name>" returns a valid IP address.
- "host <ip address>" returns a fully qualified domain name.

## Distribution server

The ADS-IL deployment scripts make use of files of files hosted on a "distribution server". During testing, this server was the hosted by James Cook University, at <http://www.hpc.jcu.edu.au/dist>. The long term future of this server is unclear.

The distribution server hosts the following files:

Directory	Contains
/pacman/VDT/vdt_181.mirror	The VDT cache, which contains code that wrappers Globus. Standard code.
/pacman/jcu/vdt_181/	Several ARCHER-developed modifications to the standard VDT cache. These add functionality such as LDAP authentication to MyProxy.
/RPM/	Yum (RPM) repository containing the <code>APAC-pacman</code> and <code>APAC-gateway-vdt-helper</code> packages. <b>Note:</b> The file <code>jcu_hpc.repo</code> in this directory should be updated to point to the location of the distribution server.

The contents of the distribution server can be downloaded from <http://www.archer.edu.au/downloads>.





## Installing the ARCHER Toolset

---

### Overview of steps

To install the entire ARCHER Toolset, proceed as follows.

Step	Approximate time
1 Decide on a configuration for the various components.	
2 (Optional) Obtain certificates and keys from your CA, if not creating a new CA.	Days or weeks.
3 (Optional) Set up a distribution host.	Several hours, including download time.
4 ADS Infrastructure Layer: Install VDT with MyProxy onto MyProxy machine.	2-3 hours
5 ADS Infrastructure Layer: Install SRB.	2-3 hours
6 Install XDMS	1-2 hours
7 ADS Service Layer: Install ICAT and MCAText.	1-2 hours
8 Install ARCHER Collaborative Workspace	2-3 hours
9 Install DIMSIM, if required.	1-2 hours
10 Install Hermes on desktop machines as required.	15 mins per machine.

The times quoted are for an experienced system administrator with knowledge of each component.

Refer to the System Administrator's Guide for each component. All documentation for each component is available at <http://www.archer.edu.au/downloads>.