caBIG_logo_w_tm

***Deployment and Administration Guide v. 0.1***

**caGRID**

**Identifier Framework**

|  |  |
| --- | --- |
| ***Contacts and Support*** | |
| Calixto Melean (Developer) | Calixto.Melean@osumc.edu |
| Scott Oster (Architect) | Scott.Oster@inventrio.com |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Revision History*** | | | |
| **Version** | **Date** | **Author** | **Changes** |
| **0.1** | 2010-07-09 | **Calixto Melean** | **Initial Draft** |

Chapter 1 Deployment Planning 5

Chapter 2 Deploying PURL 6

Installing PURLZ 6

Top Level Domain Creation 14

Chapter 3 Deploying the Naming Authority 17

Web Application 17

Grid Service 18

# Deployment Planning

Prior to deployment, the identifier prefix must be established. This involves determining the URL end points for the naming authority and the prefix authority, as well as the PURL top level domain mapping to the naming authority.

It is particularly important, and critical, to choose an appropriate domain name for the prefix authority (PURL server) and the PURL [top level domain](http://purl.oclc.org/docs/faq.html#toc1.8), as these two components make up the identifier prefix, and this is not expected to ever change after deployment. Recall that identifiers are permanent URIs by definition. The live forever.

The naming authority URL can change at anytime since it is actually hidden or protected internally by the prefix authority. When such a change occurs, the mapping from PURL domain to naming authority is simply updated to specify the new end point.

The rest of this guide assumes the following examples:

|  |
| --- |
| The Prefix Authority (PURL server) End Point |
| http://identifiers-pa.nci.nih.gov |
| PURL Top Level Domain |
| production |
| Naming Authority End Point |
| http://identifiers-na.nci.nih.gov/namingauthority/NamingAuthorityService |

With the settings above, the identifiers prefix becomes:

http://identifiers-pa.nci.nih.gov/production/

And example of identifiers are:

http://identifiers-pa.nci.nih.gov/production/030494848

http://identifiers-pa.nci.nih.gov/production/948754743

etc, etc…

When an identifier such as *http://identifiers-pa.nci.nih.gov/production/030494848* is “followed” (resolved), the prefix authority (PURL) redirects the client to *http://identifiers-na.nci.nih.gov/namingauthority/NamingAuthorityService/030494848* for resolution services.

# Deploying PURL

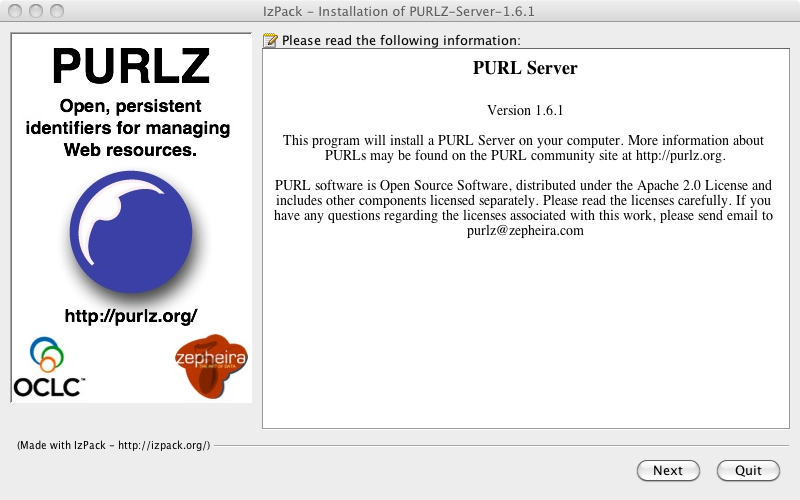
[PURLZ](http://purlz.org) is the official PURL server by OCLC. It provides a level of indirection that allows the underlying web addresses of resources to change over time without negatively affecting systems that depend on them. This capability provides continuity of references to network resources that may migrate from machine to machine.

caGrid’s identifiers framework leverages PURLZ as its prefix authority.

## Installing PURLZ

Download PURLZ: <http://purlz.org/project/purl/downloads/PURLZ-Server-1.6.1.jar>

Double click the jar file or use “*java –jar PURLZ-Server-1.6.1.jar*” from a terminal window to start the installer.



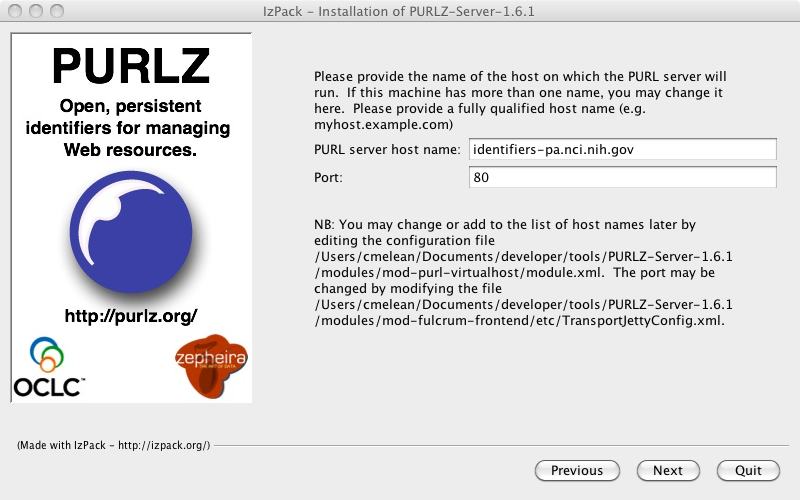
Click ***Next.***



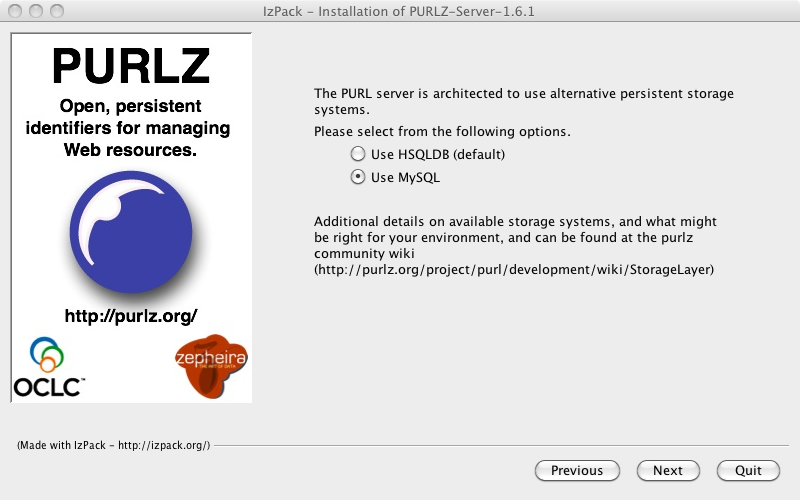
Accept the terms of the license agreement and click ***Next.***



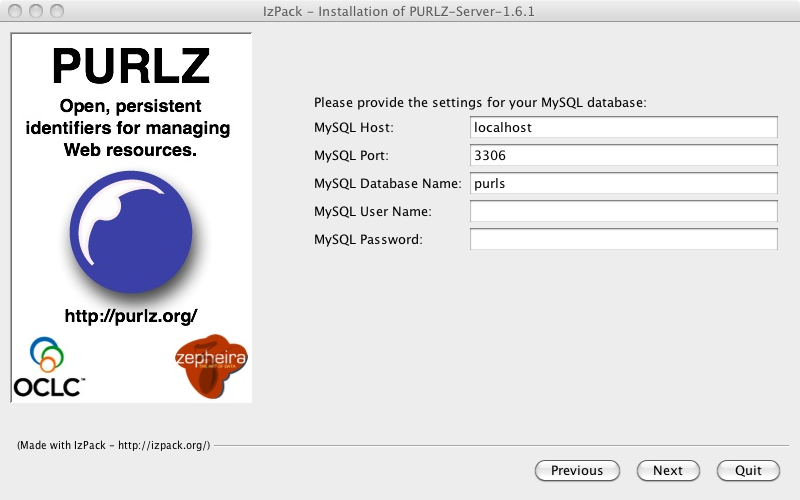
Specify an installation path and click ***Next.***



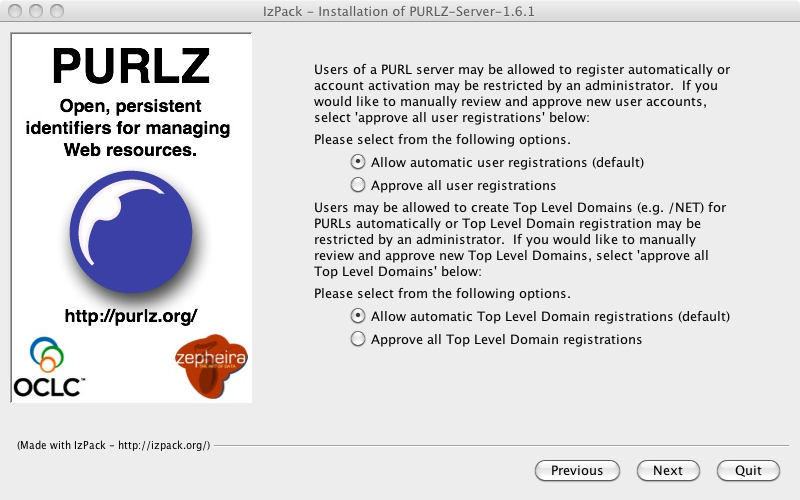
Enter the host name, port number, and click ***Next.***



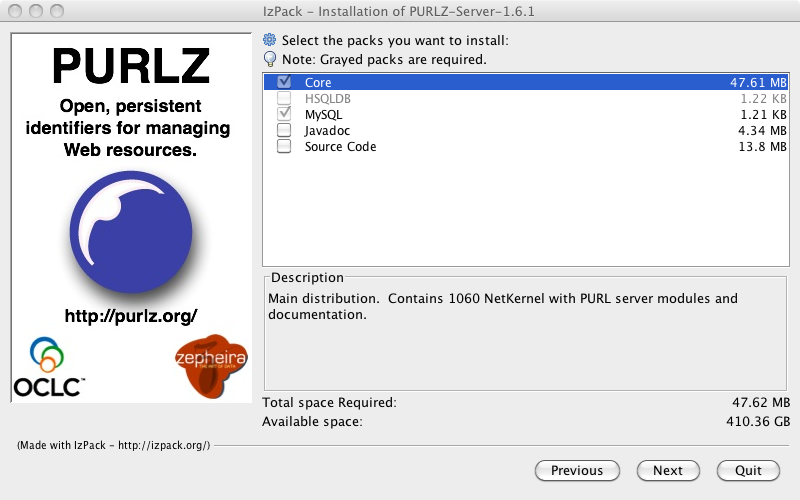
Choose “*Use MySQL*” and click ***Next.***



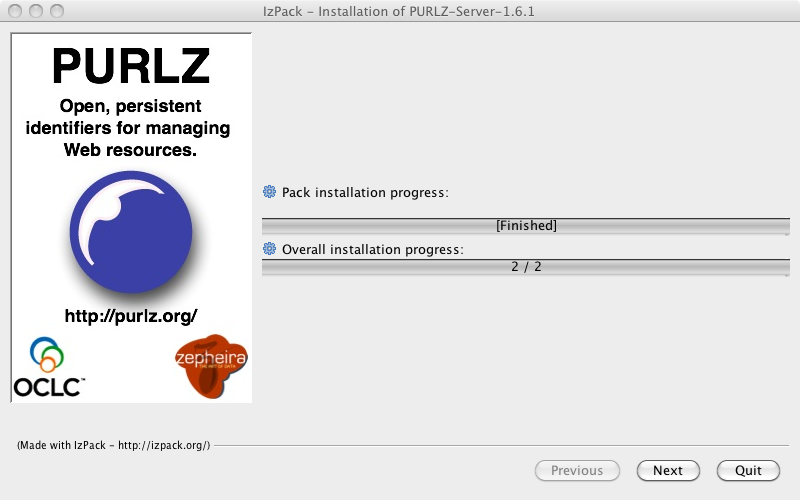
Enter *MySQL* connectivity parameters and click ***Next.***



For development and training environments, it’s probably ok to accept the defaults here. In controlled environments such as production, it is recommended that a PURL administrator be designated to approve user and top level domain registrations. Click ***Next***.



Accept defaults and click ***Next***.



The installer proceeds to complete the installation. Click ***Next*** twice and then ***Done***.

To start the server:

$ cd /Applications/PURLZ-Server-1.6.1/bin

$ ./start.sh (or startup.bat if using MS Windows)

Get started with the server by going to *http://identifiers-pa.nci.nih.gov*

You may log into the server as ‘*admin’* with the password ‘*password’* and change the password after you have logged in by modifying the user record.

## Top Level Domain Creation

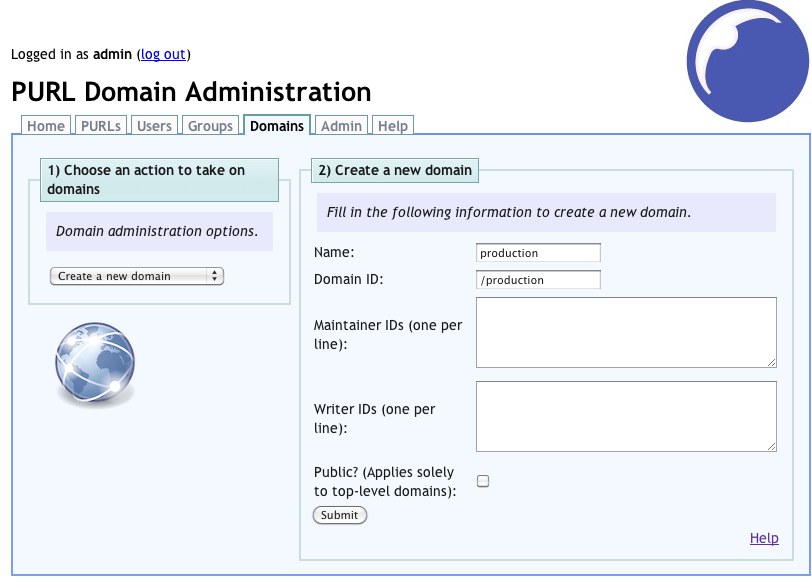
A PURL domain is needed to identify the target naming authority. The domain binds the identifier prefix to the naming authority. Therefore, a prefix authority (PURL server) can be used as an authority for multiple naming authorities by defining corresponding domains.

Following our deployment plan (Chapter 1), the goal is to create the following mapping:

***production*** => *http://identifiers-na.nci.nih.gov/namingauthority/NamingAuthorityService*

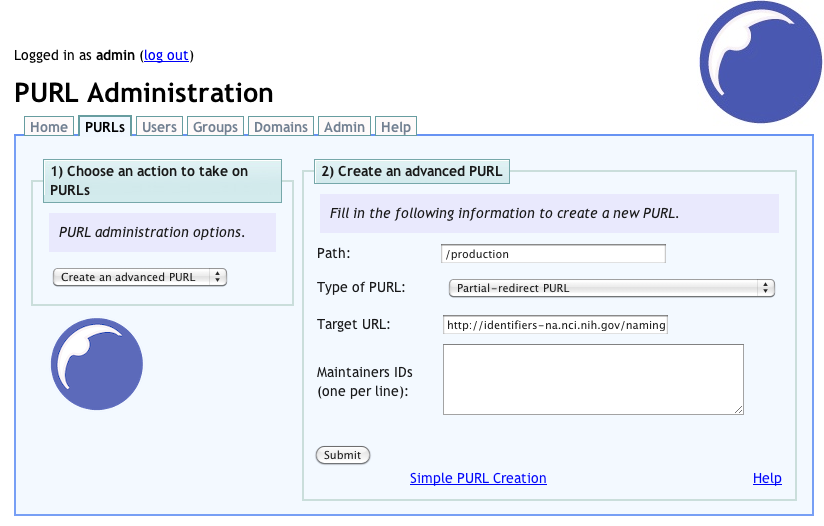
Where ***production*** is our [PURL domain](http://purl.oclc.org/docs/faq.html#toc1.8), and the mapping itself is a [partial-redirect PURL](http://purl.oclc.org/docs/faq.html#toc1.9). A domain has to be created before any PURL can placed in it.

Login as ‘*admin’* and click on *Domains* tab. Choose *Create a new domain* from the dropdown on the left, and enter the information as seen below. Click *Submit* to create the domain.



We can now create a PURL that would redirect resolution of our identifiers to their corresponding naming authority host.

Click the *PURLs* tab. Choose *Create an advanced PURL* from the dropdown on the left, and enter the information as seen below. Note that the full *Target URL* is *http://identifiers-na.nci.nih.gov/namingauthority/NamingAuthorityService*. Click *Submit* to create the PURL.



# Deploying the Naming Authority

## Web Application

This section covers the installation of the naming authority core component, the web application, which enables resolution of identifiers via HTTP.

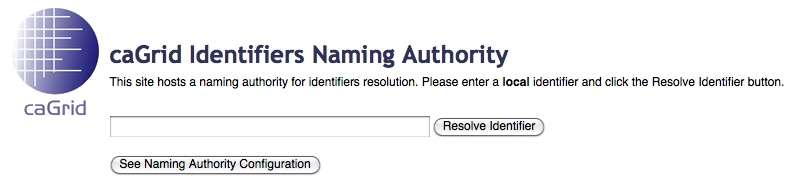
* [Install](http://cagrid.org/display/downloads/Home) caGrid 1.4 or later.
* Configure a Tomcat container to host our web application. Our deployment plan (Chapter 1) assumes this container runs on port 80.
* Customize *$CAGRID\_HOME/identifiers-namingauthority/WebContent/WEB-INF/na.properties* to match our deployment plan (*cagrid.na.prefix*). *cagrid.na.grid.url* is only needed if the optional Grid Service is being deployed.
  + **cagrid.na.prefix**=*http://identifiers-pa.nci.nih.gov/production/*
  + **cagrid.na.grid.url**=*http:*//*identifiers-na.nci.nih.gov:8080/wsrf/services/cagrid/IdentifiersNAService*
  + **cagrid.na.db.dialect**=*org.hibernate.dialect.MySQL5InnoDBDialect*
  + **cagrid.na.db.driver**=*com.mysql.jdbc.Driver*
  + **cagrid.na.db.name**=*nadb*
  + **cagrid.na.db.url**=*jdbc:mysql://localhost:3306/${cagrid.na.db.name}*
  + **cagrid.na.db.username**=*root*
  + **cagrid.na.db.password**=

**NOTE:** *cagrid.na.prefix* **MUST** include a trailing slash (/)

* Ensure the environment variable *CATALINA\_HOME* is set to point to the container previously configured.
* *cd $CAGRID\_HOME/projects/identifiers-namingauthority*
* *ant deployTomcat*
* Ensure *MySQL* is running and that the database configured above (*cagrid.na.db.name*) has been created
  + $> mysql –u root
  + mysql> create database nadb;
  + mysql> exit
* Start Tomcat
  + $CATALINA\_HOME/bin/startup.sh

(You may also want to configure Tomcat to start [automatically](http://tomcat.apache.org/tomcat-5.5-doc/setup.html)).

* Confirm the application is running by browsing *http://identifiers-na.nci.nih.gov/namingauthority*



## Grid Service

While the web application provides a *read* (HTTP resolution) interface to the naming authority, the grid service provides a *write* interface. Grid clients can leverage this interface to create and maintain identifiers.

We’ll assume here that the grid service is being deployed in its own Tomcat container running on port *8080* in host *identifiers-na.nci.nih.gov* (However, it is not necessary that the grid service be co-located with the Web Application. At minimum, at separate container is recommended).

* [Install](http://cagrid.org/display/downloads/Home) caGrid 1.4 or later.
* Configure a Globus-enabled Tomcat container (We’ll assume port 8080). This can be done with caGrid installer (previous step), or [manually](http://cagrid.org/display/knowledgebase/Manually+Configure+Tomcat+Container).
* Customize *$CAGRID\_HOME/identifiers-namingauthority-gridsvc/etc/na.properties* to match our deployment plan (*cagrid.na.prefix* and *cagrid.na.grid.url*).
  + **cagrid.na.prefix**=*http://identifiers-pa.nci.nih.gov/production/*
  + **cagrid.na.grid.url**=*http:*//*identifiers-na.nci.nih.gov:8080/wsrf/services/cagrid/IdentifiersNAService*
  + **cagrid.na.db.dialect**=*org.hibernate.dialect.MySQL5InnoDBDialect*
  + **cagrid.na.db.driver**=*com.mysql.jdbc.Driver*
  + **cagrid.na.db.name**=*nadb*
  + **cagrid.na.db.url**=*jdbc:mysql://localhost:3306/${cagrid.na.db.name}*
  + **cagrid.na.db.username**=*root*
  + **cagrid.na.db.password**=

**NOTE:** *cagrid.na.prefix* **MUST** include a trailing slash (/)

* Ensure the environment variable *CATALINA\_HOME* is set to point to the container previously configured.
* *cd $CAGRID\_HOME/projects/identifiers-namingauthority-gridsvc*
* *ant deployTomcat*
* Start Tomcat
  + $CATALINA\_HOME/bin/startup.sh

(You may also want to configure Tomcat to start [automatically](http://tomcat.apache.org/tomcat-5.5-doc/setup.html)).

* Confirm the service is running by browsing *http:*//*identifiers-na.nci.nih.gov:8080/wsrf/services/cagrid/IdentifiersNAService*



* For further verification, the *ant* target *runClient* can be run to try the creation of a simple identifier. Prior to this, ensure that the correct grid service URL is set in *run-tools.xml*



Now paste the created identifier URI into your web browser, resulting in the identifier metadata being displayed. This validates the correct operation of the prefix authority (PURL server) and the naming authority web application.

TBC