

Distributed Data Management

Jeremy Ticey, Hampton University
University of Michigan, CERN REU Program
August 14, 2008

Introduction

- ATLAS Project
- Distributed Data Management Group
- Don Quijote 2
 - DQ2 End-User Tools
- Python Programming
- Django Web Development
- Comparison/Progression

ATLAS Project

- An experiment at the Large Hadron Collider (LHC) at CERN Laboratories
- Uses high energy collisions of protons that are propelled around the radius of the accelerator
- General purpose detector that aims to study the basic forces of the universe

Group

- Distributed Data Management Group manages access to ATLAS data that is distributed to different locations around the world
- Ultimate goal is to make the search and download of datasets as simple and efficient as possible
- Numerous possibilities
 - Further development of current tools
 - Peer-To-Peer (P2P)

Don Quijote 2 (DQ2)

- Developed by the Distributed Data Management Group
- The ATLAS Experiment Data Management System
 - Bookkeeping system that stores data as datasets
- Its goal is to integrate all Grid data management services used by the ATLAS Experiment
- Provides physicists access to file-resident event data

DQ2 Clients

End-User Tools

End-User Tools

- **DQ2-Ls** -- Search for Datasets
 - [-f] -- List Files in Dataset
- **DQ2-Get** -- Download Datasets
 - [-f] -- Download a single File
- **DQ2-Put**

project goal

- To develop a prototype for a dataset search and retrieval system using a web interface

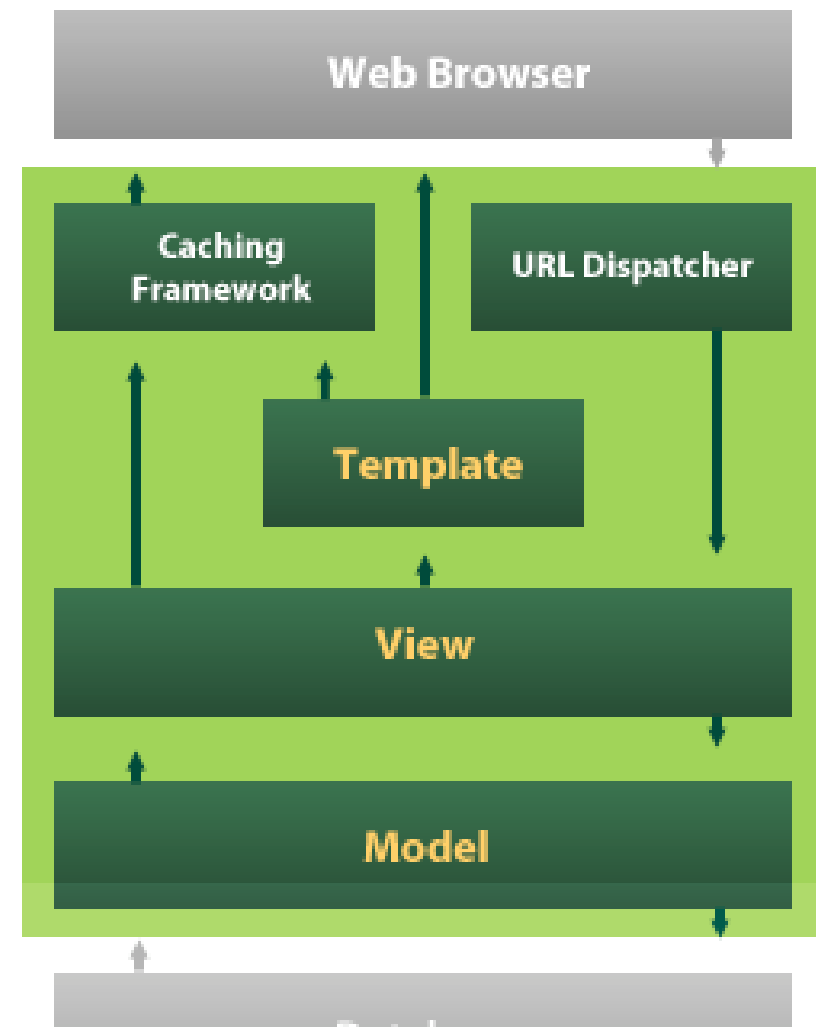


Python Script

- Python
- Developed a Dataset Search Engine
- Through terminal, you can run the program to find any dataset produced by the ATLAS experiment
- Aims to reduce the repetition of issuing commands

Django web development

- Web Framework tool that implements the Python programming language to create web pages
- Main components
 - `urls.py`
 - `views.py`
 - templates (HTML)



Comparison/ progression

- This is a prototype of one method of obtaining access to the datasets
- It will be used to determine if this is a viable option for the searching and retrieving datasets
- DQ2 -----> Python -----> Web Interface

Setup

```
*
*
ch/plus      : Information on the usage of LXPLUS/LXBATC *
ch/ComputingRules : Govern the use of CERN computing facilities *
*
*****
has been enhanced to provide an IT Service Status Board
```

...e from the IT Home Page to Useful links.

```
: http://it-div.web.cern.ch/it-div/
: http://it-support-servicestatus.web.cern.ch/it-support-servicestatus/
```

```
*****
```

```
*****
```

WARNING

```
...ly 22 2008 ATLAS Information Protection will be activated!
```

```
...at you are properly registered in the ATLAS Authorship database
```

```
...f problems mail to: atlas.info-protection@cern.ch
```

WARNING

```
*****
```

```
17]~% source /afs/cern.ch/atlas/offline/external/GRID/ddm/DQ2Clients/setup.sh
```

```
...ent AFS CERN GRID UI
```

```
...on 2.5
```

```
...ocal site
```

```
...Clients
```

```
17]~% voms-proxy-init -voms atlas
```

```
...e or dir: /afs/cern.ch/user/j/jticey/.glite/vomses
```

```
...phrase:
```

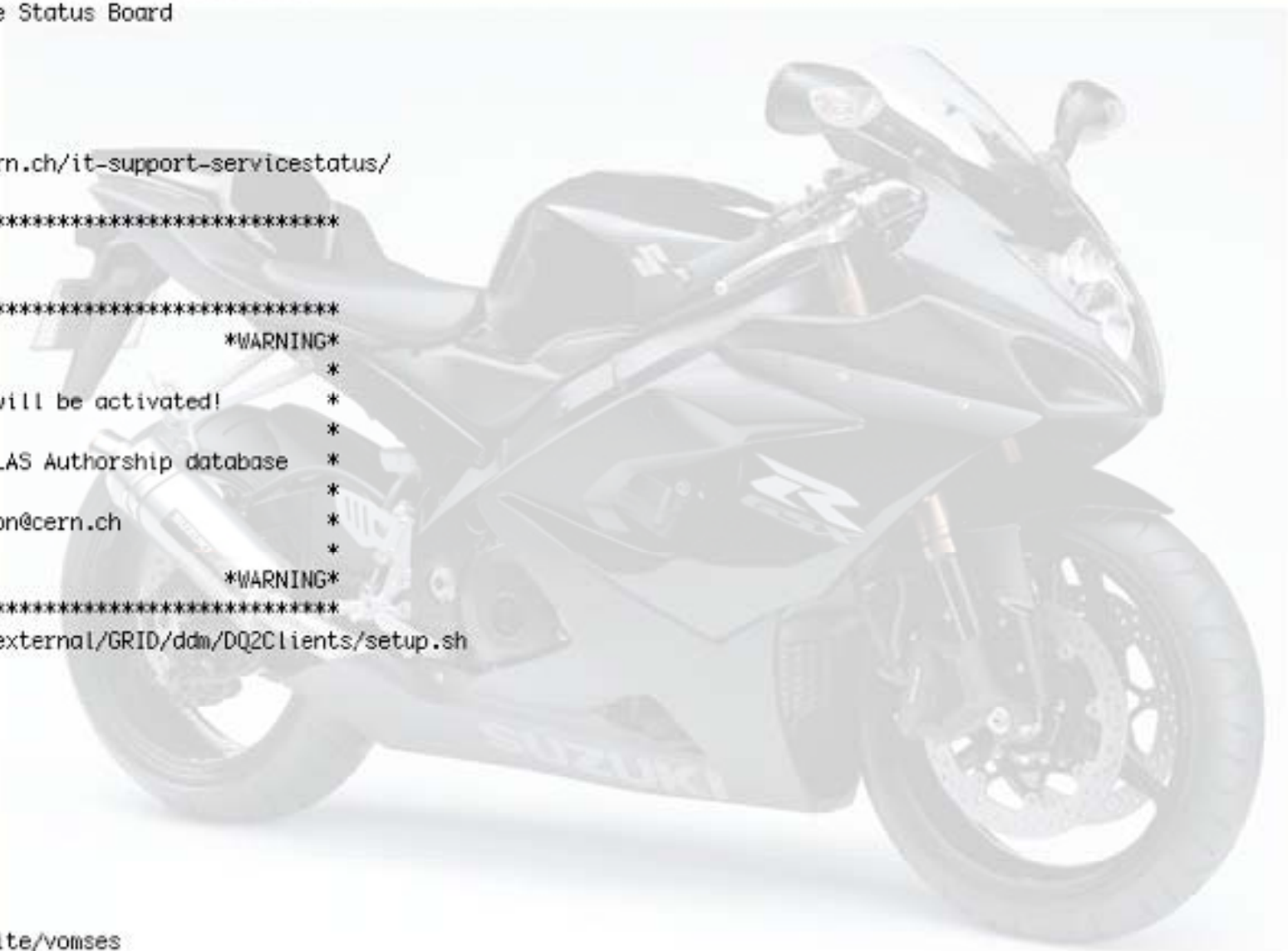
```
...se
```

```
..._init_cred
```

```
17]~% voms-proxy-init -voms atlas
```

```
...e or dir: /afs/cern.ch/user/j/jticey/.glite/vomses
```

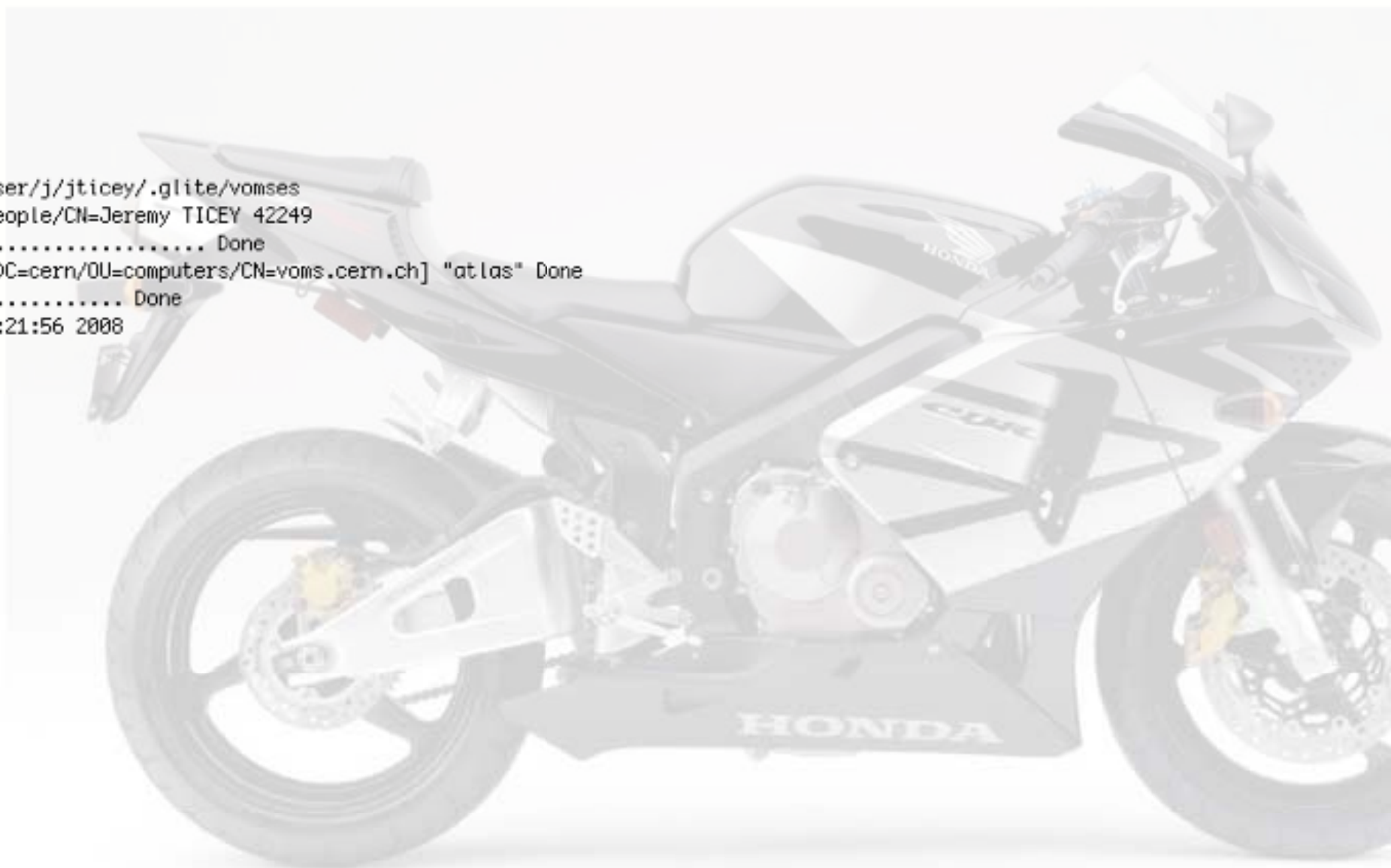
```
...phrase:
```



rent AFS CERN GRID UI
ion 2.5
local site
Clients

s phrase:

le or dir: /afs/cern.ch/user/j/jticey/.glite/vomses
/DC=org/DC=doegrids/OU=People/CN=Jeremy TICEY 42249
rary proxy Done
s.cern.ch:15001 [/DC=ch/DC=cern/OU=computers/CN=voms.cern.ch] "atlas" Done
..... Done
valid until Thu Aug 14 08:21:56 2008
are you looking for?



Sancho Panza

Search

Search

```

31]~% voms-proxy-init -voms atlas
e or dir: /afs/cern.ch/user/j/jticey/.glite/vomses
phrase:
/DC=org/DC=doegrids/OU=People/CN=Jeremy TICEY 42249
ary proxy ..... Done
-voms.cern.ch:15001 [/DC=ch/DC=cern/OU=computers/CN=lcg-voms.cern.ch] "atlas" Done
..... Done
alid until Thu Aug 14 08:24:59 2008
31]~% dq2-ls user.MiguelBranco*
found: user.MiguelBranco*
31]~% dq2-ls user.MiguelBranco
31]~% dq2-ls user.MiguelBranco.test*
found: user.MiguelBranco.test*
31]~% dq2-ls user.MiguelB*
found: user.MiguelB*
31]~% dq2-ls user.mlassnig.*
found: user.mlassnig.*
31]~% dq2-ls user.mlassnig.
31]~% dq2-ls 'user.mlassnig.*'
c3198b11f004f1193f32ab2414360ad
ernprod.001
dmd.05
374bec59bbe438fb82b21c5e73cbcad
y.happy.dataset.7
y.happy.dataset.6
y.happy.dataset.5
y.happy.dataset.4
dmd5.9
dmd5.8
estmssput.06
estmssput.07
dmd5.5
dmd5.4
dmd5.7
dmd5.6
dmd5.1
5beb24779eb4dfda172e2cd1e8faa0d
dmd5.2
2644e29d32a47c3b6f75a3ec8ff9da1
egtest.001
ataset.cern.1

```

DOUBLE-CLICK TO

ielded "173" results.

[branco.test1208433798](#)

[user.MiguelBranco.test1208422664](#)

[user.MiguelBranco.subscription.test.1208521966](#)

[user.MiguelB](#)

[branco.bookkeeping.test2.1209129190](#)

[user.MiguelBranco.test1208421436](#)

[user.MiguelBranco.test1208433870](#)

[user.MiguelB](#)

[branco.subscription.test.1208519247](#)

[user.MiguelBranco.subscription.test.1208518445](#)

[user.MiguelBranco.bookkeeping.test.1209038756](#)

[user.MiguelB](#)

[branco.subscription.test.1208518778](#)

[user.MiguelBranco.test1208424478](#)

[user.MiguelBranco.test1208423549](#)

[user.MiguelB](#)

[branco.test1208426399](#)

[user.MiguelBranco.bookkeeping.test.1208770477](#)

[user.MiguelBranco.test1208422965](#)

[user.MiguelB](#)

[branco.test1208441753](#)

[user.MiguelBranco.test1208425164](#)

[user.MiguelBranco.subscription.test.1208519406](#)

[user.MiguelB](#)

[branco.bookkeeping.test.1208519791](#)

[user.MiguelBranco.test1208433252](#)

[user.MiguelBranco.test1208443233](#)

[user.MiguelB](#)

[branco.test1208424978](#)

[user.MiguelBranco.test1208422717](#)

[user.MiguelBranco.test1208424642](#)

[user.MiguelB](#)

[branco.test1208436735](#)

[user.MiguelBranco.test1208421349](#)

[user.MiguelBranco.test1208422798](#)

[user.MiguelB](#)

[branco.test1208421427](#)

[user.MiguelBranco.test1208438541](#)

[user.MiguelBranco.test1208426078](#)

[user.MiguelB](#)

[branco.subscription.test.1208519057](#)

[user.MiguelBranco.test1208439943](#)

[user.MiguelBranco.testToday](#)

[user.MiguelB](#)

[branco.test1208425272](#)

[user.MiguelBranco.test1208423854](#)

[user.MiguelBranco.bookkeeping.test.1209106684](#)

[user.MiguelB](#)

[branco.test1208431485](#)

[user.MiguelBranco.test1208441642](#)

[user.MiguelBranco.test1208421580](#)

[user.MiguelB](#)

[branco.test1208426309](#)

[user.MiguelBranco.2](#)

[user.MiguelBranco.subscription.test.1208522306](#)

[user.MiguelB](#)

[branco.test1208432958](#)

[user.MiguelBranco.subscription.test.1208518795](#)

[user.MiguelBranco.test1208430583](#)

[user.MiguelB](#)

[branco.test1208424704](#)

[user.MiguelBranco.test1208433643](#)

[user.MiguelBranco.test1208425305](#)

[user.MiguelB](#)

[branco.mytest](#)

[user.MiguelBranco.test1208425497](#)

[user.MiguelBranco.test1208432648](#)

[user.MiguelB](#)

[branco.test1208421366](#)

[user.MiguelBranco.test1208430638](#)

[user.MiguelBranco.test1208440809](#)

[user.MiguelB](#)

[branco.test1208434026](#)

[user.MiguelBranco.test1208425367](#)

[user.MiguelBranco.mytest2](#)

[user.MiguelB](#)

[branco.test1208424309](#)

[user.MiguelBranco.testToday5](#)

[user.MiguelBranco.mytest6](#)

[user.MiguelB](#)

[branco.mytest4](#)

File Results

```
ataset.cernprodmcdisk.5
ataset.cernprodmcdisk.4
estt0.01
egtest.4
ataset.cernprodmcdisk.9
file1.37efb2aff3dc4686b3e3c9013221b904.file2.37efb2aff3dc4686b3e3c9013221b904
estmssput.05
ttest.001
67607ccf45b4954b00c82e9fd1a7d6e
8fa007720ef4da2a856ff12929010af
395332c52124d65b673f06c644b05a5
estmssput.12
estmssput.11
estmssput.10
ataset.18
ataset.19
ataset.14
ataset.15
ataset.16
estprod01
ataset.10
ataset.11
ataset.12
ataset.13
ataset.CERNPROD.2
ataset.CERNPROD.1
astor.001
astor.002
d4
atodisk.01
y.happy.tier0.dataset.01
estput.01
e966991924f462001738512301faf71
ernprod.put.002
ernprod.put.001
e0b106767164fbbb72c150df24c
fntest.04
ase.4.0
ataset.17
```

```
31]~% dq2-ls -f user.mlassnig usr.test.1
```

```
usr.test.1
lassnig usr.test.1._dummy2          b6c85475-22db-401b-8a13-bb0ded39efcf    ad:1f85c9eb    10485760
lassnig usr.test.1.dummy1          6406aa2-4444-4044-aa46-aa0404040404    ad:46a40566    10485760
```



yielded "173" results.

Phrase:

[user.MiguelBranco.test1208433798](#)

File or dir: /afs/cern.ch/user/j/jticey/.glite/vomses

/DC=org/DC=doegrids/OU=People/CN=Jeremy TICEY 42249

Primary proxy Done

g-voms.cern.ch:15001 [/DC=ch/DC=cern/OU=computers/CN=lcg-voms.cern.ch] "atlas" Done

..... Done

Valid until Thu Aug 14 08:35:09 2008

Are you looking for?

usr.test

['user.mlassnig usr.test.4', 'user.mlassnig usr.test.3', 'user.mlassnig usr.test.2', 'user.mlassnig usr.test.1']

usr.test.4

[user.MiguelBranco.test1208441753](#)

yielded "4" results.

Your search:

Are you looking for?

usr.test.1

[user.MiguelBranco.test1208424978](#)

usr.test.1

File that you were looking for?

[user.MiguelBranco.test1208436733](#)

[user.MiguelBranco.test1208421427](#)

usr.test.1

classnig usr.test.1_dummy2

classnig usr.test.1_dummy1

b6c85475-22db-401b-8a13-bbcded39efcf

ad:1f85c9eb

10485760

6d86eee3-1411-434d-aadf-ccae2de2dd36

ad:ed6e435f

10485760

[user.MiguelBranco.test1208425272](#)

071520

14 08:05:43

[user.MiguelBranco.test1208431485](#)

[user.MiguelBranco.test1208426309](#)

Like to do?

number (1-4)

[user.MiguelBranco.test1208432958](#)

the whole Dataset

part of the Dataset

[user.MiguelBranco.test1208422664](#)

[user.MiguelBranco.test1208421436](#)

[user.MiguelBranco.subscription.test.1208518445](#)

[user.MiguelBranco.test1208424478](#)

[user.MiguelBranco.bookkeeping.test.1208770477](#)

[user.MiguelBranco.test1208425164](#)

[user.MiguelBranco.test1208433252](#)

[user.MiguelBranco.test1208422717](#)

[user.MiguelBranco.test1208421349](#)

[user.MiguelBranco.test1208438541](#)

[user.MiguelBranco.test12084352943](#)

[user.MiguelBranco.test1208423854](#)

[user.MiguelBranco.test1208441642](#)

[user.MiguelBranco.2](#)

[user.MiguelBranco.subscription.test.1208518795](#)

[user.MiguelBranco.test1208433643](#)

[user.MiguelBranco.subscription.test.1208518445](#)

[user.MiguelBranco.test1208433252](#)

[user.MiguelBranco.bookkeeping.test.1208770477](#)

[user.MiguelBranco.test1208425164](#)

[user.MiguelBranco.test1208433252](#)

[user.MiguelBranco.subscription.test.1208518795](#)

[user.MiguelBranco.test1208422717](#)

[user.MiguelBranco.test1208421349](#)

[user.MiguelBranco.test1208438541](#)

[user.MiguelBranco.test12084352943](#)

[user.MiguelBranco.testToday](#)

[user.MiguelBranco.bookkeeping.test.1208770477](#)

[user.MiguelBranco.test1208426309](#)

[user.MiguelBranco.subscription.test.1208518795](#)

[user.MiguelBranco.test1208433643](#)

[user.MiguelBranco.test1208433643](#)

[6cc880e9-229a-4445-9001-6d68d41da690](#)

id: 2008-04-17 12:03:33

[c06c0563-b932-41cb-a0e3-26134794ebde](#)

id: 2008-04-17 12:03:33

download the whole dataset: [user.MiguelBranco.test1208433798](#)

Sancho Panza

[user.MiguelBranco.subscription.test.1208519057](#)

[user.MiguelBranco.test1208439943](#)

[user.MiguelBranco.test1208425272](#)

[user.MiguelBranco.test1208423854](#)

[user.MiguelBranco.test1208431485](#)

[user.MiguelBranco.test1208441642](#)

[user.MiguelBranco.test1208426309](#)

[user.MiguelBranco.2](#)

[user.MiguelBranco.test1208432958](#)

[user.MiguelBranco.subscription.test.1208518795](#)

[user.MiguelBranco.test1208424704](#)

[user.MiguelBranco.test1208433643](#)

[user.MiguelBranco.mytest](#)

[user.MiguelBranco.test1208425497](#)

[user.MiguelBranco.test1208421366](#)

[user.MiguelBranco.test1208430638](#)

[user.MiguelBranco.test1208434026](#)

[user.MiguelBranco.test1208425367](#)

[user.MiguelBranco.test1208424309](#)

[user.MiguelBranco.testToday5](#)

[user.MiguelBranco.mytest4](#)

[user.MiguelBranco.subscription.test.1208521966](#)

[user.MiguelBranco.test1208433870](#)

[user.MiguelBranco.test1208433870](#)

[user.MiguelBranco.subscription.test.1209038756](#)

[user.MiguelBranco.bookkeeping.test.1209038756](#)

[user.MiguelBranco.test1208423549](#)

[user.MiguelBranco.test1208423549](#)

[user.MiguelBranco.test1208422965](#)

[user.MiguelBranco.test1208422965](#)

[user.MiguelBranco.subscription.test.1208519406](#)

[user.MiguelBranco.subscription.test.1208519406](#)

[user.MiguelBranco.test1208443233](#)

[user.MiguelBranco.test1208443233](#)

[user.MiguelBranco.test1208424642](#)

[user.MiguelBranco.test1208424642](#)

[user.MiguelBranco.test1208422798](#)

[user.MiguelBranco.test1208422798](#)

[user.MiguelBranco.test1208426078](#)

[user.MiguelBranco.test1208426078](#)

[user.MiguelBranco.testToday](#)

[user.MiguelBranco.testToday](#)

[user.MiguelBranco.bookkeeping.test.1209106684](#)

[user.MiguelBranco.bookkeeping.test.1209106684](#)

[user.MiguelBranco.test1208421580](#)

[user.MiguelBranco.test1208421580](#)

[user.MiguelBranco.subscription.test.1208522306](#)

[user.MiguelBranco.subscription.test.1208522306](#)

[user.MiguelBranco.test1208430583](#)

[user.MiguelBranco.test1208430583](#)

[user.MiguelBranco.test1208425305](#)

[user.MiguelBranco.test1208425305](#)

[user.MiguelBranco.test1208432648](#)

[user.MiguelBranco.test1208432648](#)

[user.MiguelBranco.test1208440809](#)

[user.MiguelBranco.test1208440809](#)

[user.MiguelBranco.mytest2](#)

[user.MiguelBranco.mytest2](#)

[user.MiguelBranco.mytest6](#)

[user.MiguelBranco.mytest6](#)

[user.MiguelBranco.mytest4](#)

Download

```
...d4
datadisk.01
my.happy.tier0.dataset.01
testput.01
e966991924f462081738512301faf71
cernprod.put.002
cernprod.put.001
e0b106767164fbbb72c150df24c
fntest.04
case.4.0
dataset.17
[31]~% dq2-ls -f user.mlassnig.usr.test.1
```

```
usr.test.1
mlassnig.usr.test.1._dummy2      b6c85475-22db-401b-8a13-bbcded39efcf  ad:1f85c9eb  10485760
mlassnig.usr.test.1._dummy1      6d86eee3-1411-434d-aadf-ccae2de2dd36  ad:ed6e435f  10485760
```

```
71520
4 08:05:43
[31]~% dq2-get user.mlassnig.usr.test.1
```

```
central catalogues to resolve datasetname user.mlassnig.usr.test.1
1
usr.test.1: Querying DQ2 central catalogues for replicas...
central catalogues for files in dataset...
usr.test.1: Using site CERN-PROD_USERTAPE
usr.test.1: Querying local file catalogue of site CERN-PROD_USERTAPE...
usr.test.1/user.mlassnig.usr.test.1._dummy1: Getting metadata for srm://srm-atlas.cern.ch:8443/srm/managerv2?SFN=/castor/cern.ch/grid/atlas/tzero
mlassnig.usr.test.1._dummy1
usr.test.1/user.mlassnig.usr.test.1._dummy2: Getting metadata for srm://srm-atlas.cern.ch:8443/srm/managerv2?SFN=/castor/cern.ch/grid/atlas/tzero
mlassnig.usr.test.1._dummy2
usr.test.1/user.mlassnig.usr.test.1._dummy1: is cached at source.
usr.test.1/user.mlassnig.usr.test.1._dummy1: Starting transfer: lcg-cp -v --vo atlas srm://srm-atlas.cern.ch:8443/srm/managerv2?SFN=/castor/cern.ch/grid/atlas/tzero
usr.test.1/user.mlassnig.usr.test.1._dummy1 file:///afs/cern.ch/user/j/jticey/user.mlassnig.usr.test.1/user.mlassnig.usr.test.1._dummy1
usr.test.1/user.mlassnig.usr.test.1._dummy2: is cached at source.
usr.test.1/user.mlassnig.usr.test.1._dummy2: Starting transfer: lcg-cp -v --vo atlas srm://srm-atlas.cern.ch:8443/srm/managerv2?SFN=/castor/cern.ch/grid/atlas/tzero
usr.test.1/user.mlassnig.usr.test.1._dummy2 file:///afs/cern.ch/user/j/jticey/user.mlassnig.usr.test.1/user.mlassnig.usr.test.1._dummy2
usr.test.1/user.mlassnig.usr.test.1._dummy1: 0/10485760 transferred
usr.test.1/user.mlassnig.usr.test.1._dummy2: 0/10485760 transferred
usr.test.1/user.mlassnig.usr.test.1._dummy1: 10485760/10485760 transferred
usr.test.1/user.mlassnig.usr.test.1._dummy1: validated
usr.test.1/user.mlassnig.usr.test.1._dummy2: 3145728/10485760 transferred
usr.test.1/user.mlassnig.usr.test.1._dummy2: 4194304/10485760 transferred
```



usr.test.1
e that you were looking for?

usr.test.1
assnig.usr.test.1._dummy2
assnig.usr.test.1._dummy1

b6c85475-22db-401b-8a13-bbcded39efcf
6d86eee3-1411-434d-aadf-ccae2de2dd36

ad:1f85c9eb
ad:ed6e435f

10485760
10485760

71520
4 08:05:43

like to do?
number (1-4)

whole Dataset
part of the Dataset
another Dataset

u want to download the whole Dataset?

TLAS VOMS proxy. Create one with: "voms-proxy-init -voms atlas"

it has been downloaded to your disk. Thank you.

like to do?
number (1-4)

whole Dataset
part of the Dataset



Sancho Panza

['check.txt', 'manage.py', 'urls.py', '#downloader.py#', '# init .py#', 'DQ2Clients', 'storage', 'manage.pyc', 'pacman-3.26', 'views.pyc', 's.py~', 'check.txt~', 'user.mlassnig.usr.test.1', 'urls.pyc', 'downloader.py', 'urls.py~', 'models.pyc', 'settings.py', '#views.py#', ' init .pyc',](#)

current AFS CERN GRID UI - Enabled Python 2.5 - Configured local site - Enabled DQ2 Clients Ready Querying DQ2 central catalogue
usr.test.1 Datasets found: 1 user.mlassnig.usr.test.1: Directory user.mlassnig.usr.test.1 already exists user.mlassnig.usr.test.1: Querying
catalogues for files in dataset... user.mlassnig.usr.test.1: user.mlassnig.usr.test.1._dummy2 already transferred, validating user.mlassnig.u
r.mlassnig.usr.test.1: Using site CERN-PROD_USERTAPE user.mlassnig.usr.test.1: Querying local file catalogue of site CERN-PROD_
g.usr.test.1/user.mlassnig.usr.test.1._dummy1: Getting metadata for srm://srm-atlas.cern.ch:8443/srm/managerv2?SFN=/castor/cern.ch/g
g.usr.test.1/user.mlassnig.usr.test.1._dummy1 user.mlassnig.usr.test.1/user.mlassnig.usr.test.1._dummy1: is cached at source. user.mlassn
fer: lcg-cp -v --vo atlas srm://srm-atlas.cern.ch:8443/srm/managerv2?SFN=/castor/cern.ch/grid/atlas/tzero/atlasusertape//user/mlassnig/
g.usr.test.1._dummy1 file:///usr/local/lib/python2.5/site-packages/dq2/web/get/user.mlassnig.usr.test.1/user.mlassnig.usr.test.1._dummy1
g.usr.test.1/user.mlassnig.usr.test.1._dummy1: 0/10485760 transferred user.mlassnig.usr.test.1/user.mlassnig.usr.test.1._dummy1: 104857
g.usr.test.1/user.mlassnig.usr.test.1._dummy1: validated Finished

Conclusion

- Produced two functional dataset search and retrieval systems that will serve as prototypes for further testing
- To improve:
 - “Behind the scenes” downloader
 - Create a “waiting room” while datasets are being retrieved

Cultural Experiences

- ◆ Interlaken, Switzerland



Cultural Experiences

- ◆ Interlaken, Switzerland
- ◆ Paris, France



Cultural Experiences

Interlaken, Switzerland

Paris, France

Barcelona, Spain



Cultural Experiences

- ◆ Interlaken, Switzerland
- ◆ Paris, France
- ◆ Barcelona, Spain
- ◆ Jura Mountains



Cultural Experiences

- ◆ Interlaken, Switzerland
- ◆ Paris, France
- ◆ Barcelona, Spain
- ◆ Jura Mountains
- ◆ Geneva, Switzerland



Acknowledgements

- Miguel Branco
- Mario Lassnig
- Vincent Garonne
- Dr. Homer Neal
- Dr. Jean Krisch
- Dr. Myron Campbell
- Colleagues

Questions??

