Table of Contents

Generating a POOL XML slice using DIRAC.....1

Generating a POOL XML slice using DIRAC

The following utility has been developed in order to replace the old genCatalog functionality that was available in DIRAC2. An API function and command line script are available and allow to generate a POOL XML slice for specified LFN(s).

Please note that settings in the local configuration can affect the behaviour of this tool, for example, if the site name is not correctly specified then this must be specified on the command line. Instabilities of SRM can also give rise to problems with this tool and can be more pronounced at dCache sites where TURL construction is currently not supported.

dirac-lhcb-generate-catalog

```
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ dirac-lhcb-generate-catalog
Usage: dirac-lhcb-generate-catalog <LFN> |[<LFN>]  --site=<DIRAC Site Name> --catalog=<Catalog Fi
Try --help, -h for more information.
```

If the Site is not specified this is taken from the local DIRAC configuration e.g. /LocalSite/Site=. On lxplus shared installations it would be sufficient to specify the LFN(s) without a site name. The following optional arguments are also available on the command line:

```
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ dirac-lhcb-generate-catalog --help
2008-12-02 16:50:57 UTC Framework  INFO: Usage:
2008-12-02 16:50:57 UTC Framework  INFO: /afs/cern.ch/user/p/paterson/w1/DEV-DIRAC3/DIRAC/./LHCbS
2008-12-02 16:50:57 UTC Framework  INFO: Options:
2008-12-02 16:50:57 UTC Framework  INFO: -o:  --option=  :  Option=value to add
2008-12-02 16:50:57 UTC Framework  INFO: -s:  --section=  :  Set base section for relative parsed
2008-12-02 16:50:57 UTC Framework  INFO: -c:  --cert=  :  Use server certificate to connect to Co
2008-12-02 16:50:57 UTC Framework  INFO: -h  --help  :  Shows this help
2008-12-02 16:50:57 UTC Framework  INFO: -n:  --site=  :  DIRAC Site Name
2008-12-02 16:50:57 UTC Framework  INFO: -f:  --catalog=  :  Catalogue File Name e.g. can be /pat
2008-12-02 16:50:57 UTC Framework  INFO: -d:  --depth=  :  Optional ancestor depth to be queried
```

As an example:

dirac-lhcb-generate-catalog
LFN:/lhcb/production/DC06/phys-v2-lumi5/00001680/DST/0000/00001680_00000490_5.dst

```
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ dirac-lhcb-generate-catalog LFN:/lhcb/production/DC06/phy
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Attempting to resolve data
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Replica Lookup Time: 0.26
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Metadata Lookup Time: 0.20
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: Applying de
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: DIRAC.Workl
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: Attempting
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: InputData requi
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: 1 SURLs found f
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: CERN-disk srm:/
2008-12-02 08:51:19 UTC dirac-lhcb-generate-catalog.py  INFO: ReplicaManager.__getPhysicalFileAcc
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: Using lcg_util from: /afs/cern.ch/u
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: The version of lcg_utils is 1.6.13
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: Using gfalthr from: /afs/cern.ch/us
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: The version of gfalthr is 1.10.15
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.isValid: Determining
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.isLocalSE: Determini
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.getAccessUrl: Genera
2008-12-02 08:51:20 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.getAccessUrl: Attemp
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: Resolved input
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>> SE: CERN-d
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>LFN: /lhcb/
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>PFN: srm://
```

```
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>TURL: casto
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>PROTOCOL SR
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: All replica
2008-12-02 08:51:23 UTC dirac-lhcb-generate-catalog.py/PoolXMLSlice  INFO: POOL XML Catalogue sli
```

Since the site name and catalogue names were not specified these defaulted to LCG.CERN.ch and pool_xml_catalog.xml by default. Therefore the following XML slice was generated in the current directory.

```
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ cat pool_xml_catalog.xml
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!-- Edited By PoolXMLCatalog.py -->
<!DOCTYPE POOLFILECATALOG SYSTEM "InMemory">
<POOLFILECATALOG>


  <File ID="3E3E097D-0AC0-DB11-9C0A-00188B770645">
     <physical>
       <pfn filetype="ROOT_All" name="castor://castorlhcb.cern.ch:9002?svcClass=lhcbdata&amp;cast
     </physical>
     <logical>
       <lfn name="/lhcb/production/DC06/phys-v2-lumi5/00001680/DST/0000/00001680_00000490_5.dst"/
     </logical>
   </File>

</POOLFILECATALOG>
```

It is also possible to obtain equivalent XML slices for other sites where LHCb is able to submit data processing jobs. Looking at the available replicas of the above dataset:

```
</POOLFILECATALOG>(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ dirac-dms-lfn-replicas /lhcb/production
2008-12-02 08:52:52 UTC dirac-dms-lfn-replicas.py/DiracAPI  INFO: Replica Lookup Time: 0.21 secon
{'Failed': {},
 'Successful': {'/lhcb/production/DC06/phys-v2-lumi5/00001680/DST/0000/00001680_00000490_5.dst':
```

So another example, again from lxplus, could be to obtain an XML slice for NIKHEF in a file called myCat.xml.

dirac-lhcb-generate-catalog LFN:/lhcb/production/DC06/phys-v2-lumi5/00001680/DST/0000/00001680_00000490_5.dst --site=LCG.NIKHEF.nl --catalog=myCat.xml

```
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ dirac-lhcb-generate-catalog /lhcb/production/DC06/phys-v2
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Attempting to resolve data
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Replica Lookup Time: 0.19
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Metadata Lookup Time: 0.17
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: Applying de
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: DIRAC.Workl
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: Attempting
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: InputData requi
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: 1 SURLs found f
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: NIKHEF-disk srm
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: ReplicaManager.__getPhysicalFileAcc
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: Using lcg_util from: /afs/cern.ch/u
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: The version of lcg_utils is 1.6.13
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: Using gfalthr from: /afs/cern.ch/us
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: The version of gfalthr is 1.10.15
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.isValid: Determining
```

```
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.isLocalSE: Determini
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.getAccessUrl: Genera
2008-12-02 08:53:23 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.getAccessUrl: Attemp
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: Resolved input
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>> SE: NIKHEF
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>LFN: /lhcb/
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>PFN: srm://
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>TURL: root:
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>PROTOCOL SR
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: All replica
2008-12-02 08:53:26 UTC dirac-lhcb-generate-catalog.py/PoolXMLSlice  INFO: POOL XML Catalogue sli
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ cat myCat.xml
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!-- Edited By PoolXMLCatalog.py -->
<!DOCTYPE POOLFILECATALOG SYSTEM "InMemory">
<POOLFILECATALOG>


  <File ID="3E3E097D-0AC0-DB11-9C0A-00188B770645">
     <physical>
       <pfn filetype="ROOT_All" name="root://bee16.grid.sara.nl:1094//pnfs/grid.sara.nl/disk/lhcb
     </physical>
     <logical>
       <lfn name="/lhcb/production/DC06/phys-v2-lumi5/00001680/DST/0000/00001680_00000490_5.dst"/
     </logical>
   </File>

</POOLFILECATALOG>
```

The Catalog option also accepts full paths.

Another example using the BK ancestor depth is as follows:

dirac-lhcb-generate-catalog /lhcb/production/DC06/phys-v2-lumi2/00001758/DST/0000/00001758_00000001_5.dst --depth=1 --site=LCG.IN2P3.fr

```
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ dirac-lhcb-generate-catalog /lhcb/production/DC06/phys-v2
/afs/cern.ch/user/p/paterson/w1/DEV-DIRAC3/Linux_x86_64_glibc-2.3.4/lib/python2.4/xmllib.py:9: De
  warnings.warn("The xmllib module is obsolete.  Use xml.sax instead.", DeprecationWarning)
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Attempting to resolve data
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Replica Lookup Time: 0.18
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/DiracAPI  INFO: Metadata Lookup Time: 0.17
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: Found speci
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: DIRAC.Workl
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: Attempting
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: InputData requi
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: 1 SURLs found f
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: IN2P3-disk srm:
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: ReplicaManager.__getPhysicalFileAcc
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: Using lcg_util from: /afs/cern.ch/u
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: The version of lcg_utils is 1.6.13
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: Using gfalthr from: /afs/cern.ch/us
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: The version of gfalthr is 1.10.15
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.isValid: Determining
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.isLocalSE: Determini
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.getAccessUrl: Genera
2008-12-02 17:19:56 UTC dirac-lhcb-generate-catalog.py  INFO: StorageElement.getAccessUrl: Attemp
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: Resolved input
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>> SE: IN2P3-
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>LFN: /lhcb/
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>PFN: srm://
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>TURL: gsidc
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/InputDataByProtocol  INFO: >>>>PROTOCOL SR
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/LHCbInputDataResolution  INFO: All replica
```

```
2008-12-02 17:19:59 UTC dirac-lhcb-generate-catalog.py/PoolXMLSlice  INFO: POOL XML Catalogue sli
(paterson@lxplus244)[ ~/w1/DEV-DIRAC3]$ cat pool_xml_catalog.xml
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!-- Edited By PoolXMLCatalog.py -->
<!DOCTYPE POOLFILECATALOG SYSTEM "InMemory">
<POOLFILECATALOG>


  <File ID="8ECECC64-FDC3-DB11-BEDF-000D61CB99D2">
     <physical>
       <pfn filetype="ROOT_All" name="gsidcap://ccdcacsn008.in2p3.fr:22128//pnfs/in2p3.fr/data/lh
     </physical>
     <logical>
       <lfn name="/lhcb/production/DC06/phys-v2-lumi2/00001758/DST/0000/00001758_00000001_5.dst"/
     </logical>
   </File>

</POOLFILECATALOG>
```

For completeness, the corresponding DIRAC API function to access this functionality is:

```
from DIRAC.Interfaces.API.Dirac import *
dirac=Dirac()
dirac.getInputDataCatalog(lfns,siteName,catalogName)
```

-- StuartPaterson - 02 Dec 2008

This topic: LHCb > POOLXMLSlice
Topic revision: r4 - 2009-03-03 - unknown