

The Medical Data Management

Last review date	Reviewer
2009-09-15	Marco Bencivenni Enrico Fattibene

Table of Contents

[The Medical Data Management](#)

[Introduction](#)

[Recording a DICOM image on the grid](#)

[Retrieving a DICOM slice](#)

[Retrieving a 3D image](#)

[Removing an image](#)

[How to give permissions to another user](#)

The Medical Data Management

How to use the Medical Data Manager command- line to create, manage and remove medical images and metadata.

Introduction

The Medical Data Manager (MDM) is an interface between DICOM (Digital Image and COmmunication in Medicine) compliant storage and the gLite middleware. It aims at (i) providing access to medical data sources for computing without interfering with clinical practice, (ii) ensuring transparency so that accessing medical data does not require any specific user intervention, and (iii) ensuring a high data protection level to preserve patients privacy.

This service exploits the DICOM standard for medical image transfers on the clinical side and the Storage Resource Management (SRM) standard for grids. It bridges these two standards by translating on the fly grid file read accesses into DICOM transactions. It benefits from the EGEE middleware capability to manage distributed files, thus enabling the federation of many DICOM servers geographically distributed and it provides a unified view of the data archived. It exploits state of the art encryption and fine grain ACL- based mechanisms to ensure both data protection and access control.

This document first describes the MDM middleware functionality and architecture. It then describes five use cases that illustrate most of the MDM command lines. You can find more details on other command lines and configuration files [on the MDM web site](#). Some parts of this document refer to outside sections. You can find their references in the [???](#) page.

Description of the MDM middleware

The MDM is a distributed collection of services, built on top of the EGEE Data Management System, aiming at securely managing medical image file stores using the native DICOM protocol on the grid. This section describes this middleware. The use cases illustrate how the services are inter- operated.

- **DICOM server:** A DICOM server is a server which stores medical images in a DICOM format, and uses the DICOM image communication protocol. The MDM interfaces to a single DICOM server or a DICOM protocol- compliant PACS (hospital Picture Archiving and Communication System). The MDM is meant for interfacing to an existing, pre- deployed clinical PACS. For testing purposes the Conquest DICOM server is provided with the MDM.
- **DPM: Disk Pool Manager.** DPM provides an SRM (Storage Resource Manager) compliant interface to disk storage hosted on regular workstation(s) (single disk or disk pool). A DPM plug- in has been developed to support a DICOM server back- end. The SRM storage space is used to cache the image files retrieved.
- **DCMTK:** DCMTK is a collection of libraries and applications implementing large parts of the DICOM standard. It includes software for examining, constructing and converting DICOM image files, handling offline media, sending and receiving images over a network connection, as well as demonstrative image storage and work- list servers. DCMTK is written in a mixture of ANSI C and C + +. It comes in complete source code and is made available as "open source" software. The DPM- DICOM plug- in uses this library for DICOM data manipulation.

- **AMGA: ARDA Meta Catalog.** AMGA is a grid-enabled front-end for various relational database back-ends. The MDM uses AMGA to store medical metadata associated to the images manipulated. The authentication is based on grid credentials. The AMGA schema associates grid file LFNs (Logical File Name) to the DICOM identifier of a medical image. The [AMGA section](#) describes in detail the AMGA metadata catalog.
- **Hydra:** The Hydra service is an encrypted storage solution. It enables encryption of the files that are stored on storage resources. The sensitive information is the encryption key, which is stored in the Hydra Keystore and access controlled. Hydra splits and distributes the keys to several (e.g. three) keystores. The keys shares are partially redundant (e.g. 2 out of 3 shares are needed to reconstruct an encryption key) and always incomplete (e.g. at least 2 shares are needed). Thus even if one server fails or is compromised, the keys are still accessible and protected (see Shamir's Secret Sharing Scheme).
- **LFC: LCG File Catalog.** The LFC provides a unified hierarchical view of files distributed over storage resources. The user-defined virtual path of a file is named Logical File Name (LFN). This path is associated to one (or several in case of replication) physical file location(s). The user can access a file without knowing where it is physically stored. The [Basic Data Management section](#) gives more details on the LFC and the file data management system.

DICOM standard and grid filename

DICOM is the most established standard for medical data management. Digital Imaging and Communications in Medicine (DICOM) covers both a medical image data format and an image communication protocol. A DICOM image usually contains one slice (a 2D image) acquired using any medical imaging modality (MRI, CT-scan, PET, SPECT, ultrasound, X-ray...). A DICOM image contains both the image data itself and a set of metadata related to the image, the patient, the acquisition parameters and the radiology department. DICOM metadata are stored in fields. Each field is identified by a unique tag defined in the DICOM standard. The Medical Data Manager reads these tags to create a database for the medical information. Each DICOM file is unique and three DICOM fields in each slice identifying the study, exam, and slice number. These fields are named STUDY Instance UID (for Unique Identifier), SERIES Instance UID and SOP Instance UID. The three numbers are used by the MDM to identify each image. The MDM registers each image in the grid catalog. The grid filename is `/grid/ <vo>/mdm/ <STUDY_Instance_UID>/ <SERIES_Instance_UID>/ <SOP_Instance_UID>`.

In practice, during a medical examination (a study in the medical world), a radiologist acquires several 3D images, representing up to hundreds to thousands of slices. A study is composed by one or several examination. The MDM allows the user to retrieve all the slices of an examination in a 3D file. The grid filename of the 3D images is `/grid/ <vo>/mdm/ <STUDY_Instance_UID>/ <SERIES_Instance_UID>/3D`. The grid filenames are named Logical File Name (LFN). You should read the [Basic Data Management section](#) for more details.

Security

On a grid, the distribution of data makes security a very sensitive problem. The Medical Data Manager checks the user permissions to retrieve a DICOM file at multiple levels. To ensure patient privacy, the header of all DICOM images sent by a DICOM server should be deleted, at least partially, to ensure anonymity. After the anonymization and before sending, the images are encrypted. The encryption keys are stored in the hydra keystore service. The keys are associated to the image LFN. The MDM hides the security mechanisms providing a unique command line that retrieves and deciphers the 3D image or DICOM file.

Certificate and proxy

To access the grid and the MDM the user first has to create a proxy. The proxy contains his credentials and provides single sign-on for all the EGEE services. The proxy is based on the user certificate. The [Proxy Certificates section](#) describes in details the proxy and the [Getting a Certificate section](#) explains the certificate and the way to get one.

Each use cases needs a valid proxy, but the proxy initialization is not specific to the Medical Data Management. It will not appear in the use cases and this section gives only basic information about it.

How to initialize a proxy:

```
voms-proxy-init -voms <vo>
```

This command asks the user to give the password for his certificate. An example:

```
voms-proxy-init -voms biomed
```

The output is different for each user, but there is much similarity:

```
Cannot find file or dir:
/home/texier/.glite/vomses Enter GRID pass phrase: Your identity:
/O=GRID-FR/C=FR/O=CNRS/OU=I3S/CN=Romain Texier Creating temporary
proxy ..... Done Contacting
cclcgvomsl01.in2p3.fr:15000
[/O=GRID-FR/C=FR/O=CNRS/OU=CC-LYON/CN=cclcgvomsl01.in2p3.fr]
"biomed" Done Creating proxy
.....
Done Your proxy is valid until Fri Apr 17 03:56:59 2009
```

Recording a DICOM image on the grid

To register a DICOM image in the Medical Data Manager, the user should use the following command line:

```
glite-mdm-register [--verbose/-v] [--very-verbose/-vv] <local DICOM file>
```

The mandatory parameter is the pathname of the DICOM image (for example: /home/user/MedicalImage/DicomImage.dcm). The two other parameters define the level of verbosity. Without any verbosity option, there is no output in case of success. This command line will also create a 3D image. All the 2D DICOM image of the same patient, study and series will be gather to create the 3D image.

Example:

```
glite-mdm-register /home/user/MedicalImage/DicomImage.dcm
```

There is not output, but you can query the LFC or the AMGA server to see the new DICOM image.

Internal behavior

A DICOM images recording is a multi step process. The first steps concern the DICOM slices recording and subsequent steps deal with the 3D image. The 3D image is a collection of slices that does not correspond to additional entries in the DICOM server but that needs to be identified in the file catalog.

- The glite-mdm-register first step is to create an encryption key in the Hydra server. This key will be used to cipher and decipher the DICOM image. No raw DICOM files are sent to the grid network.
- The second step records the DICOM file in the DICOM server. This step use the DCMTK library to communicate with the DICOM server.
- The third step create an entry in the DPM server. A DPM entry is called physical file name (PFN). The MDM adds special PFNs, because the file is not stored inside the DPM. The PFN refers to the DICOM data that is handled by the DPM- DICOM plug-in.
- The fourth step register the DICOM file in the LFC. The filename template is /grid/"vo"/mdm/"STUDY Instance UID"/"SERIES Instance UID"/"SOP Instance UID". The MDM creates the directory for the first image.
- The next step stores the metadata of the medical image in the AMGA server. The metadata describes the images details and some patient information.
- The following steps record the 3D image. The DPM- DICOM plug-in retrieves all the 2D slices from the DICOM server and it creates the 3D image.
- The next step is to register a new entry in the DPM server.
- The last step registers the 3D image in the LFC catalog. The grid filename is /grid/"vo"/mdm/"STUDY Instance UID"/"SERIES Instance UID"/3D.

Retrieving a DICOM slice

To retrieve a DICOM image from the Medical Data Manager, there are two steps. First, the AMGA Metadata Catalog is used to find the image identifier. Then the `glite-mdm-get` command line is used to retrieve the DICOM slice.

To access the AMGA Metadata Catalog, the user must launch the multi-server AMGA client. This client will allow the user to connect to AMGA servers, which contain medical metadata. By default, only the local AMGA server is used. The AMGA client allows the user to make SQL-like requests. This section gives only one request that retrieved all the image identifiers of a patient. More information is available in [AMGA section](#).

```
mmdclient selectattr /mdm/IMAGE:FILE '/mdm/PATIENT:name="patient name" and \
/mdm/PATIENT:patientID=/mdm/IMAGE:patientUID'
```

The first command line connects to the default server, that gives a greeting.

```
mmdclient
Connected to localhost:8822 : ARDA Metadata Server 1.3.0
Query>
```

The following query returns the Globally Unique Identifier (GUID) of the DICOM slices.

```
selectattr /mdm/IMAGE:FILE '/mdm/PATIENT:name="Romain Texier" and \
/mdm/PATIENT:patientID=/mdm/IMAGE:patientUID'
>> guid:00ac24cb-5dc4-4784-a7f2-d10215b01895
>> guid:ba9d3e44-6e0d-41f3-8db3-36511071e5fb
>> guid:0971f517-c1c9-48a0-ae04-a5bceb054702
>> guid:a85c56f0-ae70-497e-870c-5dc628e5676c
>> guid:9a16cc3c-da0a-4f00-9e69-59dead1a07aa
>> guid:97873ef6-e324-4d1f-a8cf-9f2b0bc2d002
>> guid:11303975-97bd-48bf-aa21-da3d9cc9d0ea
>> guid:5aba484e-9115-49ee-9638-e5be2be255ec
```

The `glite-mdm-get` retrieves a DICOM picture from an MDM server and deciphers it. There are four ways to specify the DICOM slice. All of them are equivalent. If the destination file doesn't exist, the command will be cancelled (see `-i` and `-f` options). If the verbose option is not enabled, there is not output on success.

```
glite-mdm-get -s/--soap <study>,<serie>,<sop> <destination filename>
glite-mdm-get [-s/--soap] [/"study"/<serie"/<sop> <destination filename>
glite-mdm-get [-l/--lfn] <lfn> <destination filename>
glite-mdm-get [-g/--guid] <guid> <destination filename>
```

with the options

```
-i/--interactive, -f/--force, -v/--verbose
```

The meaning of the arguments:

`<destination filename>`

"destination filename" is a local filename where the retrieved file will be stored. For example : / home/ user/ medical- picture/ new-dicom.dcm)

`<study>,<serie>,<sop>`

"study","serie","sop" is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a space or comma between values.

`[/"study"/"serie"/"sop"`

`[/"study"/"serie"/"sop"` is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a slash "/" between the values and an optional slash at the beginning.

The meaning of the options:

-v or --verbose

this software will be verbose.

-i or --interactive

this software will ask to confirm. It prevents, for example, to overwrite a file.

-f or --force

this software will overwrite the destination file without any notification.

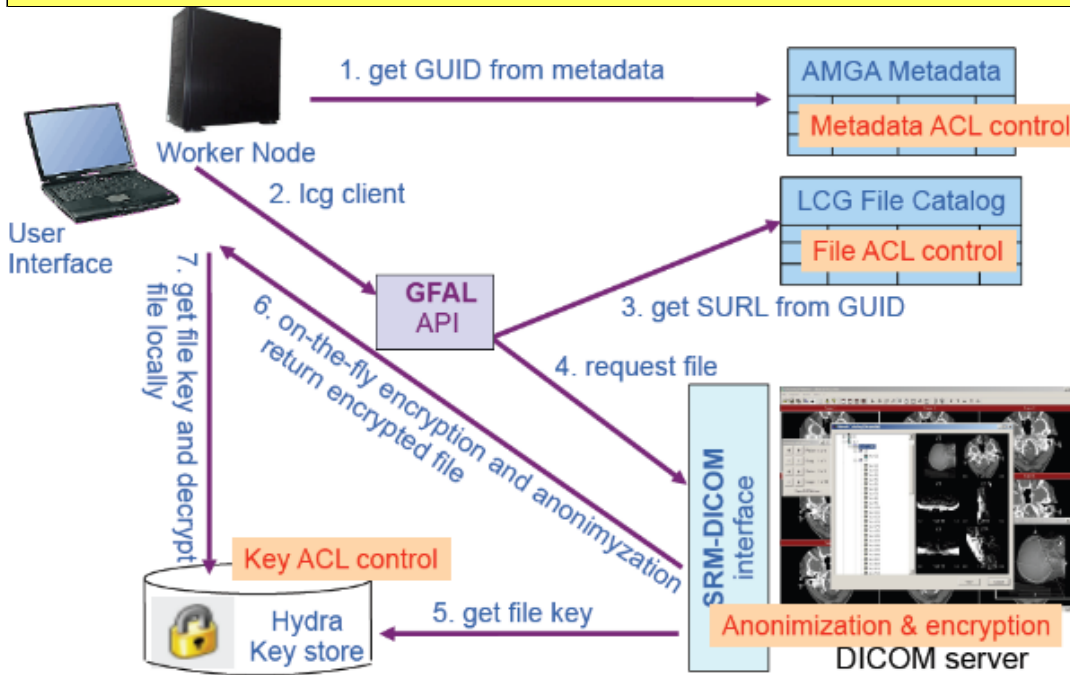
Example:

In this example, the GUID of the DICOM picture is used:

```
glite-mdm-get guid:a85c56f0-ae70-497e-870c-5dc628e5676c /home/user/<DestinationDicomFile>
```

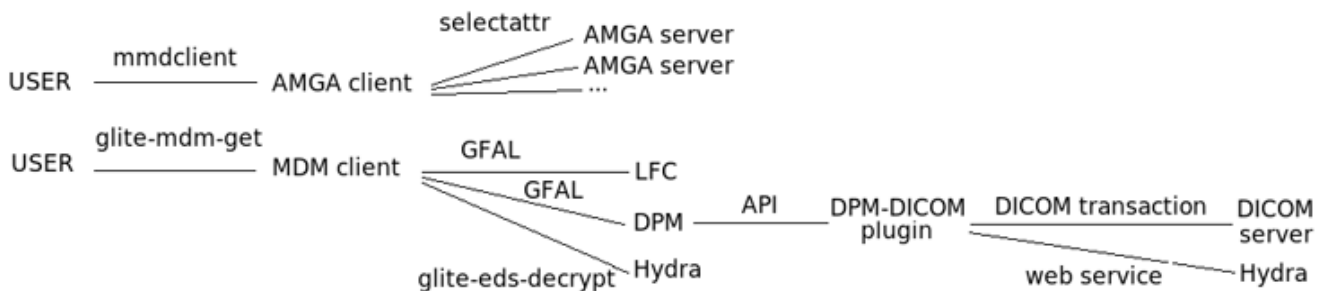
Usual glite command lines can also be used to retrieve a DICOM image. In this case decryption needs to be performed manually through the glite-eds-decrypt command.

```
lcg-cp <LFN> <filename-encrypted>
glite-eds-decrypt </study/serie/sop> <filename-encrypted> <filename>
```



Retrieving a DICOM slice: The services called

The main services, which are used to retried a DICOM slice, are described in this picture:



Retrieving a 3D image

The way to retrieve a 3D image from the Medical Data Manager is similar to the DICOM slice case. The AMGA Metadata Catalog is used to find the image identifier. Then the `glite- mdm- get` command line is used to retrieve the 3D image.

To access the AMGA Metadata Catalog, the user must launch the multi- server AMGA client. This client will allow the user to connect to AMGA servers, which contain medical metadata. By default, only the local AMGA server is used. The AMGA client enable the user to make SQL- like requests. This section gives only one request that retrieved all the image identifiers of a patient. More information is available in [AMGA section](#).

```
mmdclient selectattr
/mdm/PATIENT:name /mdm/STUDY/:FILE /mdm/SERIE/:FILE
'/mdm/PATIENT:name="patient name" and /mdm/PATIENT:FILE=/mdm/STUDY:MDMpatientID and \
/mdm/STUDY:FILE=/mdm/SERIE:STUinsUID'
```

The first command line connect to the default server, that gives a greeting.

```
mmdclient
Connected to localhost:8822 : ARDA Metadata Server 1.3.0
Query>
```

The following command line asks for all the 3D image identifiers of a patient. The AMGA servers send back the studies and series instances of the 3D images.

```
selectattr /mdm/PATIENT:name /mdm/STUDY/:FILE /mdm/SERIE/:FILE '/mdm/PATIENT:name="Romain Texier" \
and /mdm/PATIENT:FILE=/mdm/STUDY:MDMpatientID and /mdm/STUDY:FILE=/mdm/SERIE:STUinsUID'
```

In this example, there are two 3D images. The first image has the same value for its study instance and its serie instance.

```
>> Romain Texier
>> 1.2.826.0.1.3680043.2.1143.74144555039.20060227135016132.72
>> 1.2.826.0.1.3680043.2.1143.74144555039.20060227135016132.72
>> Romain Texier
>> 1.2.826.0.1.3680043.2.1143.74144555039.20060227138345524.45
>> 1.2.826.0.1.3680043.2.1143.74144555039.20060227135044253.87
```

The `glite- mdm- get` command retrieves medical images from an MDM server and deciphers them. The command line parameters are similar to the 2D use case. The 3D image does not have an instance number. This number is replaced by the value "3D". There are four way to specify the 3D image. All of them are equivalent. If the destination file exists, the command will be cancelled (see `-i` and `-f` options). If the verbose option is not enabled, there is no output on success.

```
glite-mdm-get -s/--soap "study","serie",3D "destination filename"
glite-mdm-get [-s/--soap] [/"study"/"serie"/3D "destination filename"
glite-mdm-get [-l/--lfn] "lfn" "destination filename"
glite-mdm-get [-g/--guid] "guid" "destination filename"
```

with the options

```
-i/--interactive, -f/--force, -v/--verbose
```

"destination filename"

"destination filename" is a local filename where the retrieved file will be stored. For example : / home/ user/ medical- picture/ new- dicom.dcm)

"study","serie","sop"

"study","serie" is the Study Instance UID and the Serie Instance UID with a space or comma between values.

[/"study"/"serie"

[/"study"/"serie" is the Study Instance UID and the Serie Instance UID with a slash "/" between the values and an optional slash at the beginning.

The meaning of the options :

- v or -- verbose

this software will be verbose.

- i or -- interactive

this software will ask to confirm. It will prevent to overwrite a file.

- f or -- force

this software will overwrite the destination file without any notification.

Example:

In this example, the instance values find in the AMGA server are used.

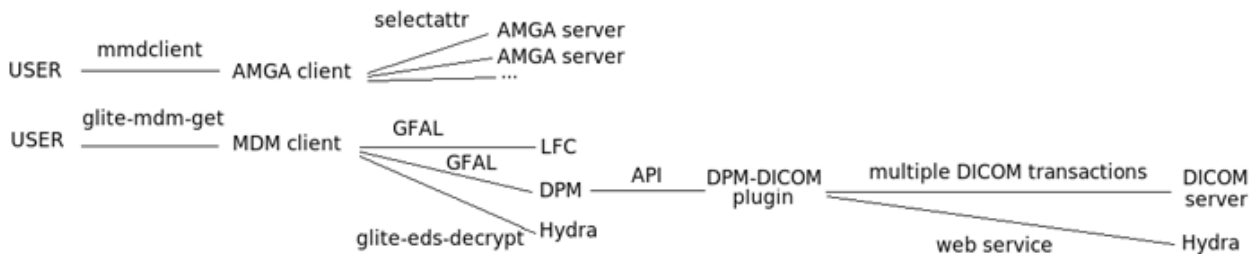
```
glite-mdm-get --sop 1.2.826.0.1.3680043.2.1143.74144555039.20060227135016132.72/  
1.2.826.0.1.3680043.2.1143.74144555039.20060227135016132.72/3D/home/user/<DestinationDicomFile>
```

Usual gLite command lines can also be used to retrieve a DICOM image. In this case decryption needs to be performed manually through the `glite-eds-decrypt` command.

```
lcg-cp <LFN> <filename-encrypted>  
glite-eds-decrypt </study/serie/sop> <filename-encrypted> <filename>
```

Retrieving a DICOM slice: The services called

The main services, which are used to retried a DICOM slice, are described in this picture:



The schema is similar to the previous use case schema. The difference is the multiple DICOM transactions between the DPM- DICOM plugin and the DICOM server.

Removing an image

A DICOM image can be removed from the Medical Data Manager. The DICOM image will not be anymore available for grid user, but the image will remain in the DICOM server. The DICOM server specific interface should be used to remove the image from the DICOM server. To remove the DICOM image from the grid, you need an identifier of this image. There are multiple ways to identify the DICOM image. These four ways are equivalents.

```
glite-mdm-del [-l/--lfn] "lfn"  
glite-mdm-del -s/--soap "study","serie","sop"  
glite-mdm-del -s/--soap [/"study"/"serie"/"sop"
```

with the option

```
-v/--verbose
```

"lfn"

The Logical File Name (LFN) of the image. This identifier is used by the [Lcg File Catalog \(LFC\)](#).

"study","serie","sop"

"study","serie","sop" is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a space or comma between values.

[/]"study"/"serie"/"sop"

[/]"study"/"serie"/"sop" is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a slash "/" between the values and an optional slash at the beginning.

Example:

In this example, the LFN value is used to remove the DICOM slice.

```
glite-mdm-del --lfn /grid/biomed/mdm/1.2.826.0.1.3680043.2.1143.74144555039.20060227134827193.29/1.2.826.0.1.3680043.2.1143.74144555039.20060227134827193.29/\1.2.826.0.1.3680043.2.1143.74144555039.20060227134827510.76
```

Removing a DICOM slice: The services called

The main services, which are used to remove a 3D image, are described in this picture:



How to give permissions to another user

Any owner of DICOM images can share them with others. There are multiple levels of sharing. The user can share images with users, groups of users or everyone. An user, who is allowed to access an image, can retrieve it with the `glite-mdm-get` command line. The use case number two describes how to do it. The owner can also share the medical metadata. The MDM split these metadata in two parts. These two parts are the patient nominative data and the anonymous image information. These two kinds of metadata can be shared independently. An user can give access to the anonymous information and he can simultaneously hide the patient metadata. All the metadata are registered in AMGA servers. In the servers, there are default groups: Register, Physician, MIA, biomed. The "register" users are the only users, who can add new metadata. "Physician" are the users, who are authorized to read the patient metadata. MIA stands for Medical Image Analysist. This group of people may be allowed to read part of the metadata. The "biomed" users are the members of the owner VO. Members may have access to the anonymous metadata.

```
glite-mdm-set-right [-l/--lfn] "lfn" [rwxrwxrwx]
                    [private|group|anonymous|public]
                    [ -u|--user "DN" "rwx" "none|anonymous|all" ]
glite-mdm-set-right [-f/--file] "local DICOM image" [rwxrwxrwx]
                    [private|group|anonymous|public]
                    [ -u|--user "DN" "rwx" "none|anonymous|all" ]
glite-mdm-set-right -s/--soap "study","serie","sop" [rwxrwxrwx]
                    [private|group|anonymous|public]
                    [ -u|--user "DN" "rwx" "none|anonymous|all" ]
glite-mdm-set-right -s/--soap [/"study"/"serie"/"sop" [rwxrwxrwx]
                    [private|group|anonymous|public]
                    [ -u|--user "DN" "rwx" "none|anonymous|all" ]
```

with the option

```
-v/--verbose
```

The first part of the command line is similar to previous use cases. It selects the image:

"lfn"

The Logical File Name (LFN) of the image. This identifier is used by the [Lcg File Catalog \(LFC\)](#).

"local DICOM image"

"local DICOM image" is a copy of the image on the local file system.

"study","serie","sop"

"study","serie","sop" is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a space or comma between values.

[/]"study"/"serie"/"sop"

[/]"study"/"serie"/"sop" is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a slash "/" between the values and an optional slash at the beginning.

"study","serie","sop"

"study","serie","sop" is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a space or comma between values.

[/]"study"/"serie"/"sop"

[/]"study"/"serie"/"sop" is the Study Instance UID, the Serie Instance UID and the SOP Instance UID with a slash "/" between the values and an optional slash at the beginning.

The second part of this command line is specific. It defines the permissions to apply.

"rwxrwxrwx"

"rwxrwxrwx" is the global permissions for the image. Those permissions are similar to the Unix permissions. The letters mean read (r), write (w), list (x). The "x" permission allow to list the DICOM image. The "r" permission allow to get the DICOM image. To retrieve a file, the "x" and "r" must be set. The "w" permission allow to change the image for another one. There are very few reasons to give this permission. The first three letters are the permission for the owner of the image. The second "rwx" is the VO user permissions. The last "rwx" is for people of other VO. The "-" character means that the permission is denied. The permissions "rwxr-x---" gives all permissions to the owner and these permissions allow users of the same VO to retrieve the image.

private|group|anonymous|public

private|group|anonymous|public are the metadata permissions. In all cases, the owner and the "register" group can read and write the metadata. The private mode does not allow anyone else to access the metadata. The public mode allows the four groups to access all the metadata. The biomed users and the MIA can read the anonymous medata in the "anonymous" mode. The "group" mode give different permissions to each group. In the "group" mode, the physician can read metadata. The MIA can access anonymous information and the biomed users can not read any metadata.

-u|--user DN rwx none|anonymous|all

The -u|--user "DN" "rwx" "none|anonymous|all" option gives or removes permissions for a specific user. The Distinguish Name is the unique name of this grid user. The letters rwx give the permissions on this image. The "none" permission denies access to the metadata. The "anonymous" permission allows the user to read the anonymous metadata. The "all" permission give access to all the metadata of this image.

Example:

In this example, the instance values are used to set restrictive permissions on the DICOM file.

```
glite-mdm-set-right rwxr----- group --sop \  
/1.2.826.0.1.3680043.2.1143.74144555039.20060227134724311.31/\   
1.2.826.0.1.3680043.2.1143.74144555039.20060227134724311.31/\   
1.2.826.0.1.3680043.2.1143.74144555039.20060227134725065.30
```

Give permissions : The services called

The main services, which are used to edit the permissions, are described in this picture:

