Create and update Rucio Datasets from DUST

After creating files on the DUST scratch disk (i.e., somewhere under `/nfs/dust/atlas/`), you might want to upload it to ATLAS' Rucio namespace, so that colleagues at other institues can copy/replicate your data.

Prerequisites

On a NAF workgroup server, you will need the ATLAS environment to be setup including the rucio client

```
> export ATLAS_LOCAL_ROOT_BASE=/cvmfs/atlas.cern.ch/repo/ATLASLocalRootBase
> source ${ATLAS_LOCAL_ROOT_BASE}/user/atlasLocalSetup.sh
> lsetup rucio
```

and to upload your data, you also need a valid grid proxy

```
> voms-proxy-init -voms atlas
```

Creating a dataset

If you have not have already a Rucio dataset, i.e., a container for your files, you can create a new one with something like

```
> rucio add-dataset user.thartman:mytestdataset
```

with the syntax user.<YOURATLASNAME>:<NEWDATASETNAME>

Uploading Files to a grid storage

As the DUST space is only available internally to DESY/NAF-users, we have to upload files to a Rucio Storage Element in the Grid.
To upload a file called "`ubuntu_sandbox.tar.xz`" to my Rucio scope "`user.thartman`"

```
> rucio upload ubuntu_sandbox.tar.xz --scope user.thartman --rse DESY-HH_SCRATCHDISK
```

here we used the DESY-HH_SCRATCHDISK storage element to send the file to, because it is close by in the network and is a scratchdisk to transient data. For a list of all available storage elements check out

```
> rucio list-rses
```

Please ask your group and the ATLAS data management team, what storage to use - especially if you plan to use a lot of data or want to keep your data over longer time periods.
Else you might get angry emails fast.

Attaching a file to a dataset

The uploaded file is now on the choosen Rucio Storage Element, but not yet attached to the previously created dataset. But we might want to collect a number of files in one dataset, so that we can easily ship the whole collection within the ATLAS collaboration with replication rules etc. .

So, we attach the file to the dataset

```
> rucio attach user.thartman:mytestdataset user.thartman:ubuntu_sandbox.tar.xz
```

and can further operate with the dataset as adding further files or setting up replication rules to other Rucio Storage Elements.

Uploading a directory and registering it directly to a dataset

The separate steps to create a dataset, upload a file and attach the file to a dataset can be tedious.
Fortunately, `rucio uploads` supports also a bulk upload with attaching the files to a dataset; unfortunately the syntax is a bit cumbersome:

```
> rucio upload --rse DESY-HH_SCRATCHDISK --scope user.thartman user.thartman:anotherdataset MYDIRECTORY.d/
```

would upload files in a directory `MYDIRECTORY.d/` to my scope `user.thartman` and register it in a dataset user.thartman:anotherdataset.
Please note, that this is not recursive - only files in `MYDIRECTORY.d/` will be uploaded but no files in sub-directories of `MYDIRECTORY.d/`

Links

- https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/RucioClientsHowTo