**Open Exeter project, University of Exeter, UK**

DSpace comes with its own ‘out of the box’ submission tool which works fine with small files in small numbers using real time transfer over http. What if I am a researcher with datasets whose file sizes are of the order of TBs – I really can’t wait that long and don’t want to depend on someone from the IT department manually importing these for me !

The UK JISC funded Open Exeter project set out to understand how researchers at the University of Exeter manage their data.

As part of the project researchers were surveyed about the amount of data they stored and how they stored it particularly once a project was finished. It was found that some research projects produced huge numbers of files some of which were individually very large and that these were often archived on local hard disks and external drives not necessarily on the local network. To submit these datasets to our DSpace based institutional repository is not practical using the out of the box DSpace submission tool since it limits the user to one file at a time for upload over HTTP whilst the user waits.  In addition transferring large files via such methods can be slow and prone to failure. DSpace also supports command line batch import of files providing they can be successfully transferred via some other means. SWORD provides a standardised way of interfacing with repositories including DSpace but also currently remains limited in its ability to transfer large files reliably.

To overcome these limitations Open Exeter has developed its own submission tool using elements of the SWORD protocol combined with the leading research data transfer service Globus Online. SWORD allows us to query the repository to determine which collection the user is allowed to submit to and what sort of metadata is needed. The tool then gathers the metadata and data locations from the user, creates a metadata only “placeholder” deposit via SWORD and starts transfer of the files in the background using Globus Online to facilitate the transfer. It is important to note that the data itself does not flow through Globus but is securely sent peer to peer.

The user disconnects from the tool and a background service uses Globus monitoring capabilities to determine when the file transfer is complete. At this point the placeholder deposit is revisited and completed by using the DSpace “import by reference” built in functionality using a custom SWORD artefact to support this action specifically for DSpace. The user is emailed with a unique reference when the placeholder submission is accepted and again when the submission is finalised. Normal approval workflow rules apply if they are turned on.

This method works irrespective of the volume of data and its location whilst remaining secure. Added benefits to using Globus include faster transfers using multiple channels for both single or multiple file sets, interruption recovery and the ability to work across network boundaries (no firewall issues !).

The first time a user accesses the service they are guided through the process of creating a Globus account and associating this with their University credentials. They may also be asked to create a Globus Connect “endpoint” if this is the first time their data location has been used for transfer. Subsequent access is seamless and uses Single Sign On technology to securely authenticate the individual user with DSpace and Globus.

Using these technologies, Open Exeter is working towards a solution that will allow researchers to upload their data quickly and securely. We will be delighted to provide a live demonstration of the service and take any questions or comments you may have.