**Summary**

With the introduction of Rose/cylc, post-processing and archiving of model data is handled by the post-processing app. UM fields files can be converted to PP, NetCDF files compressed and rebuilt and the usual meaning performed. Data is then moved to an archive.

The use of a scheduler (cylc) to manage experimental workflows means that data can be processed and moved as the simulation is running, greatly improving overall efficiency.

NCAS-CMS has extended the Met Office’s post-processing system (which archived to MASS only) to cover our multi-platform research environment, specifically:

- Archiving to the Research Data Facility (RDF) from ARCHER
- Addition of a separate task to transfer the archived data to JASMIN

**Data Workflow**

![Data Workflow Diagram]

**Data Transfer**

After data has been archived to the RDF (or a staging area on NEXCS), a suite can be configured to automatically transfer the data to JASMIN.

Data can be transferred using either rsync or GridFTP and is configurable through the Rose GUI. For high volumes of data GridFTP, configured to use multiple streams, to the high-speed data transfer node at JASMIN (jasmin-xfer2) should be used.

Once the files have been transferred, checksums are used to verify that the transfer was successful. Checksum verification takes a significant proportion of the whole transfer task time, but is essential to be sure data has not been corrupted.

<table>
<thead>
<tr>
<th>Approximate Data Transfer Rates in Mb/s</th>
<th>jasmin-xfer1</th>
<th>jasmin-xfer2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Push from RDF:</td>
<td>15</td>
<td>127</td>
</tr>
<tr>
<td>rsync</td>
<td>46</td>
<td>150</td>
</tr>
<tr>
<td>GridFTP</td>
<td>-</td>
<td>20*</td>
</tr>
</tbody>
</table>

* Upgrade to faster network planned for late January will improve rate

**Performance**

Performance of the post-processing task has improved significantly since the early releases. On ARCHER some of the high resolution GA configurations were taking longer to post-process and archive the data than it had taken the model to run.

Performance has been improved by:

- Using different conversion utilities
- Ability to split atmos, nemo and cice post-processing into separate tasks thus running them in parallel
- Option to further split processing and archive tasks. This would allow intensive processing to be run on a different machine, potentially parallel nodes, to the archiving.

**Availability**

The post-processing, archive and transfer functionality for ARCHER / NEXCS has been incorporated into the Met Office MOCI trunk and is available at release postproc_2.2. Instructions on how to upgrade and setup the postproc app in your suite and how to transfer data to JASMIN can be found on the CMS website:

http://cms.ncas.ac.uk/wiki/Docs/PostProcessingApp

**Future Work**

- Use of GridFTP to transfer data from NEXCS to JASMIN
- Archive Integrity app – check for gaps in the RDF archive