



# DiRAC deploys Atempo Miria for Archiving

CLIENT: DIRAC/UK

SECTOR: ADVANCED RESEARCH/HPC



Four UK universities, Cambridge, Durham, Edinburgh & Leicester, are responsible for delivering DiRAC's HPC services.

These universities provide core HPC facilities along with expertise to enable multiple research, support, knowledge transfer and industrial partnership projects. The DiRAC Project Office is hosted by University College London.

Significant data considerations are at the heart of DiRAC's remit; petaflop compute and petabyte storage requirements are integral to DiRAC-supported projects.

---

*Significant data considerations are at the heart of DiRAC's remit; petaflop compute and petabyte storage requirements are integral to DiRAC-supported projects.*

## About DiRAC

DiRAC (Distributed Research Utilizing Advanced Computing) is the integrated supercomputing facility for theoretical modeling and HPC-based research in particle physics, astronomy and cosmology and nuclear physics. It is a key part of the infrastructure supporting the UK's Science and Technology Facilities Council (STFC) Frontier Science programme. Research scientists across the UK can apply for access to DiRAC's powerful computing facilities.

## DiRAC and Atempo: some background

Recently, DiRAC's Memory Intensive facility in Durham called on the services of Atempo, the Data Protection and Movement specialists, together with their UK partner, OCF, to implement a multi-petabyte archiving project for their Lustre and Spectrum Scale (GPFS) data.

We spoke to Dr Lydia Heck, DiRAC Technical Manager and Dr Alastair Basden, Technical Lead for the DiRAC Memory Intensive service, both of whom are based in Durham University's Institute for Computational Cosmology (ICC).

The DiRAC Memory Intensive service, the seventh increment in a series of HPC clusters at Durham, provides researchers with **452 nodes and 12,656 cores of computing power**.

Lydia and Alastair gave us their feedback on Atempo's Miria for Archiving solution which meets the requirements for research project archiving to tape. We also received some insight into how archiving and backup needs may evolve in the coming months.



One of DiRAC@Durham's recent missions was to switch from Spectrum Scale (GPFS) storage to DDN's Lustre storage. The aim being, in the words of Lydia, *"to effectively enrich storage with a less expensive solution."* The growing need for ever-higher memory intensive computing generates significant data volumes within HPC environments.

Additional compute power is closely aligned with increased data storage requirements and greater tiering needs. Generated data storage by 2022 is projected to reach **upwards of 20 petabytes** with the roll out of DiRAC Phase 3. DiRAC's Data Management Plan includes an archival component for both the research database and finished peer-reviewed scientific research documents. It is the research data which requires archiving.

## The solution in practice: Atempo Miria for Archiving

The role of Miria is to archive research data and relieve higher cost disk storage by offloading DiRAC research data from primary storage to four LTO tape destinations. Current storage requirements stand at 6.5 PBs with an uplift to 20 PBs when DiRAC Phase 3 is rolled out (expected early 2020) which will see a **10-fold increase in processing** and a corresponding augmentation in data creation and therefore storage requirements.

The incumbent archiving solution was slow and not as scalable as required. Atempo proposed a Proof of Concept (POC) in late 2018 on Spectrum Scale GPFS first and then on Lustre. **Atempo rapidly demonstrated that Miria could make short work of DiRAC's data archiving needs.** A Miria Archiving Server and a dedicated Miria Data Mover directly access Lustre file systems and efficiently handle all data archive flows from source to destination.

The Miria for Archiving User Interface **allows users to perform their own rapid archiving tasks** and to restore their data process. Researchers can use the logical file tree format of their choice for each research project and preserve this tree structure when archiving their data. The Administration interface enables DiRAC to manage both archiving and backup of critical data.



*"Miria for Archiving is incredibly powerful and feature-rich and should meet our future needs even though we've barely scratched the surface with the solution, we're very impressed so far."*

*Dr Alastair Basden,  
Technical Lead for the  
DiRAC Memory Intensive  
Service,  
Durham University*

update 2019-05-27



The DiRAC technical teams at Durham are also using Miria to back up user home space along with archiving files selected by users. *“Every administrator and user action is handled through the HTTP protocol which means it is very easy for us to set up an SSH tunnel and enable users to archive their files wherever they are in the world.”* All the physical data moving equipment along with the Miria server is installed and running and Alastair reports that: *“archiving performance on Lustre file systems data flows is running at full tape speeds, which is perfect.”*

Archiving requirements include creating file data copies on two tape locations. The DiRAC service at Durham is using LTO7m media today with plans to bring in LTO8 tapes when these become readily available. Atempo fully supports this technology mix.

## Looking ahead

Miria is not just an archiving tool. The Miria for Backup component also protects user pool data. During the POC Atempo also **demonstrated FastScan capabilities for GPFS**. FastScan optimizes how new and modified files are recognized and stored to avoid lengthy file system rescanning. A key future integration for DiRAC will be to leverage Miria for its advanced backup capabilities with Atempo’s future FastScan capacity for DDN Lustre.

The strength of Atempo’s R&D teams lies in their in-depth industry knowledge of file systems, handling data attributes and striping options for example. DiRAC are keen to work alongside the DDN and Atempo teams for this integration and provide access to DiRAC’s valuable beta testing on an HPC-scale file and storage environment.

“This is exactly what DiRAC expects from its active technology providers. To help justify expenses in significant resources, we need to be able to tell our stakeholders that we are also on the forefront of new and innovative solutions in many domains which include data movement and data protection,” Lydia concludes.

DiRAC has several highly connected geographical sites between the member Universities, reaching **bandwidths of 20Gbit/s**. DiRAC will consider future projects that move, protect and store data off site. Making a secondary data copy of one site’s data and storing this data on a different site is one potential evolution DiRAC will evaluate.

To conclude, the Miria Archiving solution is running with very positive results in terms of performance, efficiency and security. The Atempo and DiRAC teams will continue to work in a spirit of cooperation and mutualized resources to build on this initial success.

