TELOS Data Management update

Ed Bennett

on behalf of the TELOS Collaboration

LDG meeting, Odense, 2025-04-04

Background

- Aims: study theories of relevance to Beyond the Standard Model physics
  - SU(2) with adjoint matter
  - Sp(4) with fundamental and antisymmetric matter
- Codes:
  - HiRep (including modifications for Sp(2N) theories, currently in fork)
  - Grid (including modifications for Sp(2N) theories, merged)

Ensemble status

Group	Fermion Action	N_{f}	N_{as}	N_{adj}	N_{PV}	V_{max}	β	Status	Storage
SU(2)	Wilson			1		96×48^3	$[2.05, 2.4]$	Ready	36TiB
SU(2)	Wilson			2		128×64^3	$[2.25,2.35]$	Ready	16TiB
SU(2)	Wilson			1, 2	5, 10, 15	40^4	$[2.35, 2.7]$	In preparation	O(15TiB)
SU(2)	Möbius			1, 2	tbc	tbc	tbc	In preparation	
Sp(4)	Wilson	2				48×42^3	$[6.9, 7.5]$	Ready	24TiB
Sp(4)	Wilson		3			54×36^3	$[6.6, 6.9]$	Ready	16TiB
Sp(4)	Wilson	2	3			56×36^3	$[6.45, 6.5]$	Ready	28TiB
Sp(4)	Möbius	2				tbc	tbc	In preparation	

Data readiness

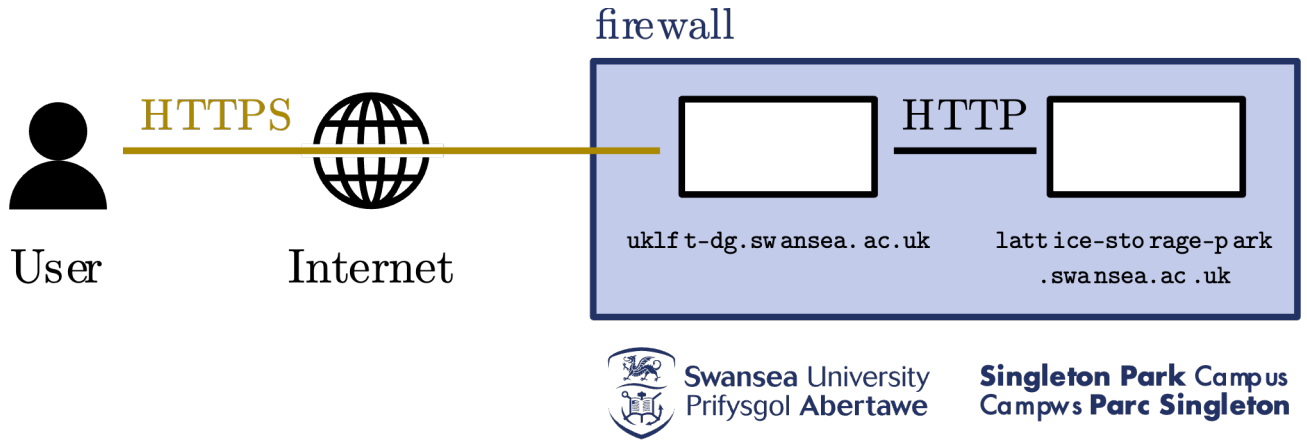
- Awaiting availability of UK metadata and file catalogues
- Remaining to do:
  - Convert configurations to ILDG format
    - Making use of new reduced storage format
  - Generate and upload ILDG metadata for ensembles and configurations

Data storage

- 400TB storage server in Swansea shared with FASTSUM collaboration
  - Funded from STFC Consolidated Grant
  - Aiming to expand with next Consolidated Grant round
  - Challenging to ask for more than £10k
- Ongoing discussions around DiRAC Data Curation Service
  - Aim to have long-term storage for data generated by projects on DiRAC HPC services

Data access

- Experimenting with proxied token-authenticated HTTPS access to storage
  - Swansea University is selling IP space
  - Work is on hold until addressing stabilises



- DiRAC Data Curation Service *should* support Globus protocols

Non-configuration data

- Share on Zenodo:
  - **Post-HPC data** for re-analysis and reproducibility
    - E.g. HMC logs, Correlation functions, Gradient flow histories
    - Raw logs and/or HDF5 format
  - **Post-analysis data** for ease of quoting results
    - E.g. Spectroscopic values, fit parameters
    - CSV and/or SQLite format
  - **Analysis and data presentation workflow** for reproducibility and understanding
    - Aim for one-click end-to-end analysis reproducibility
- Issues:
  - Zenodo record size limit
  - Standardisation of data and metadata formats