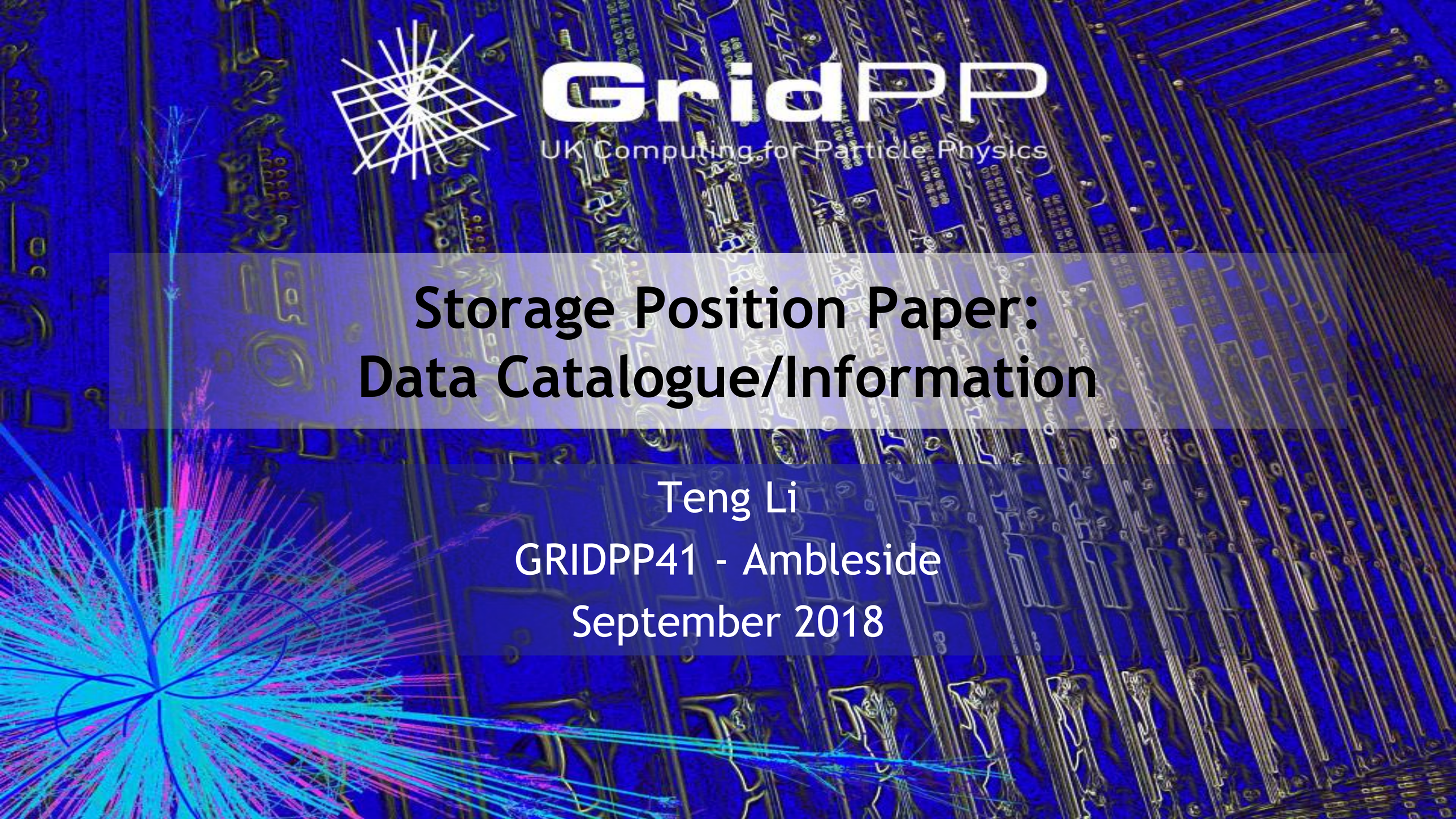

Storage Position Paper: Data Catalogue/Information

Teng Li

GRIDPP41 - Ambleside

September 2018

GridPP
UK Computing for Particle Physics

Contents

- Data Catalogue
  - Current Status
  - Future Developments
  - GridPP's Role
- Information Management
- Summary
- Discussion Reportage at Meeting

Current Status

- LCG File Catalog
  - Going away
  - Smaller VOs still using
    - Need migration plan
- CMS TFC
  - Scalability Concerns

Current Status

- DIRAC (DFC):
  - LHCb remains the heaviest user
  - In wide use but limited to experiments which have integrated their workflow with DIRAC
  - Within GridPP, DIRAC instance at Imperial supports ~17 small VOs (lz, pheno, snoplus, gridpp, lsst, …)
  - Non-HEP (and even some nonWLCG HEP) experiments often use their own DIRAC file catalogue at Imperial - is this sufficient? Rucio replacement of DFC within DIRAC stack?

Current Status

- Rucio:
  - Current users: ATLAS, ASGC: AMS + others, Xenon1T
  - Within HEP:
    - Becoming more widely adopted with CMS looking to migrate their data management before the HL-LHC
    - DUNE is also experimenting Rucio
  - New and existing astronomy VOs are looking to use Rucio for data management/cataloguing
    - SKA
    - OSG (LIGO, IceCube)
    - EISCAT_3D
    - …
  - In one word, rucio is becoming widely used

Data Catalogue/Management

- Future Development
  - Existing data management solutions will likely require some development to remove any dependency on GridFTP/SRM protocols in the near future.
  - Distributed resilient storage (c.f Data Lakes) will require changes to the concept of the data catalogue.
  - Rucio, and potentially other data management tools, are adding representations for volatile data (inspectable caches) in their catalogues.
- What needs doing
  - Consolidation.
  - Move away from proprietary databases.

Data Catalogue/Management

- GridPP's Role
  - Give advices for new VOs
  - Implementation/incorporation of existing VOs data models.
  - Flexibility to change if/when data model changes.

Information System

- Accounting, APEL
  - To understand where future accounting is going. Know what is coming up in changes to APEL
  - Is it worth working much on APEL given that LHC VOs are doing their own thing?
- BDII
  - Should we/How should we get rid of BDII
  - How much this would affect SEs?
- Resource cataloguing by VO.
  - AGIS / TFC + site_config.xml + Phedex / Alien / DIRAC