

# DOMA project status

Simone Campana

Maria Girone

(CERN)

# The WLCG Data Organization Management Access evolution project (DOMA)

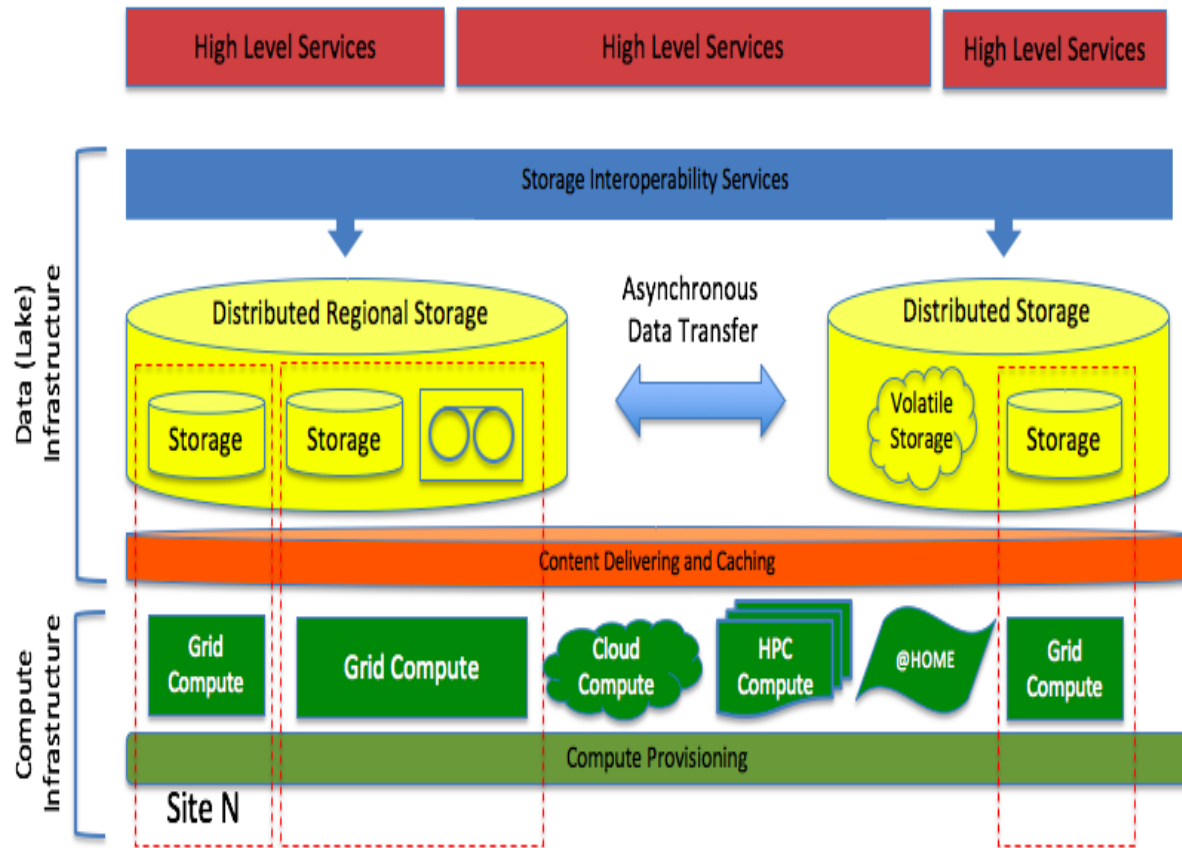
A set of R&D activities evaluating data management solutions for HL-LHC, based on the expected evolution of facilities

- Fewer and larger storage endpoints
- Heterogeneous CPU resources
- Data/processors co-location not guaranteed

The ultimate goal is to meet the computing needs at affordable cost

Many already existing efforts. DOMA provide a forum to discuss ideas and foster interoperability of solutions

- an umbrella for experiments, middleware developers and storage providers, facilities



# DOMA info

- Meeting every 4<sup>th</sup> Wednesday of each month, from 16:00 (finishing before 17:30)
- One topical discussion at every meeting + short report for each working group
- DOMA general Mailing List: `wlwg-doma (at) cern.ch`
- Three active working groups:
  - Third Party Copy (TPC) protocols: `wlwg-doma-tpc (at) cern.ch`
  - Data Access, Content Delivery and Caching (ACCESS): `wlwg-doma-access (at) cern.ch`
  - Storage Quality of Service (QoS): `wlwg-doma-qos (at) cern.ch`
- DOMA twiki: <https://twiki.cern.ch/twiki/bin/view/LCG/DomaActivities>
  - The twikis of the working groups are all linked from here

# Third Party Copy

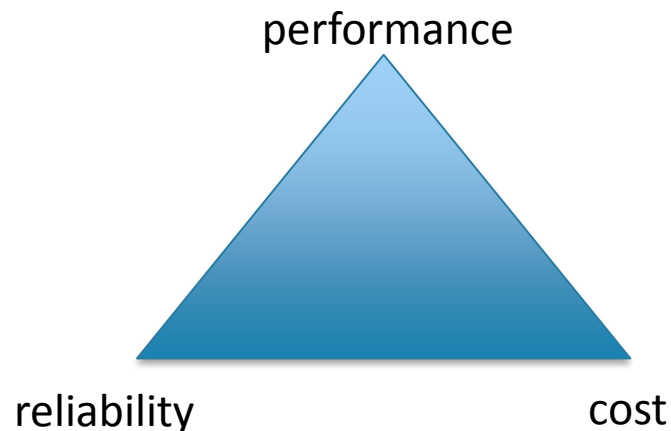
- Chairs: Alessandra Forti and Brian Bockelman
- <https://twiki.cern.ch/twiki/bin/view/LCG/ThirdPartyCopy>
- Short term goal: investigate, commission and deploy alternative TPC protocols to gridFTP
  - Three phases (milestones) finishing in Dec 2019 with all sites providing storage to WLCG offering a non-gridFTP endpoint
- Medium term goal: prototype token-based auth in TPC
  - Focus on bearer tokens (capability based): Macaroons and SciTokens
  - In line with the WLCG AAI task force (see later)
- Xrootd and HTTP/Dav are the candidate protocols

# ACCESS

- A broad scope WG: data access performance, content delivery and caching
- Chairing team: S. Jezequel, I. Vukotic, F. Wuerthwein, X. Espinal, M. Schulz
- <https://twiki.cern.ch/twiki/bin/view/LCG/ContentDeliveryCaching>
- Started by looking at existing activities in this broad domain and organize the information by topic
  - [https://docs.google.com/document/d/1Sk5wtFLdHDCjyc\\_VmTwJY4qzk\\_ErKidBs7GLhTnN4xo/edit](https://docs.google.com/document/d/1Sk5wtFLdHDCjyc_VmTwJY4qzk_ErKidBs7GLhTnN4xo/edit)
- Two main topics:
  - Data access patterns and access performance studies
  - Organizing and deploying caching solutions

# Quality of Service

- Chaired by Paul Millar and Oliver Keeble
- The high level goals are:
  - At the storage level, define, implement and expose different classes based on performance/reliability need and what you can afford
  - Integrate the notion of storage classes in the higher levels, such as experiment DDM systems. Leveraging on existing efforts such as XDC
  - Synergies with the WLCG archival storage working group



- Not a new concept: What we call “Disk” and “Tape” are in fact QoS
- A potential source of large hardware saving for HL-LHC and favors the integration of new storage technologies
- Trade off performance, reliability and cost based on the use case

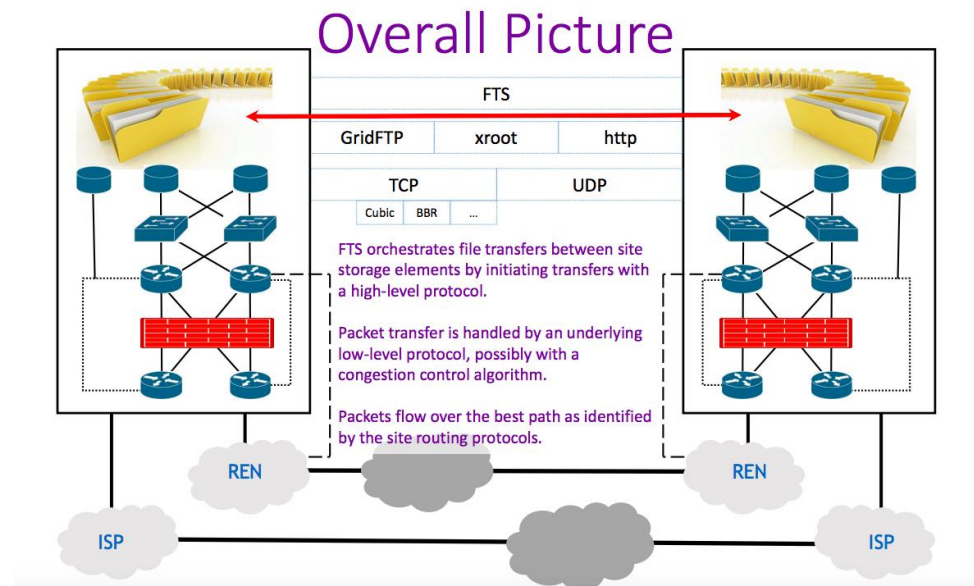
# DOMA related network activities

- Network R&D activities, focusing on data transfer
  - DTNs, low level transfer protocols, bandwidth on demand, P2P channels, SDNs, ...

- Collaboration with the SKA AENEAS project and HEPIX

- Leveraging information from FTS as file transfer manager

- <http://cern.ch/go/7qxT>



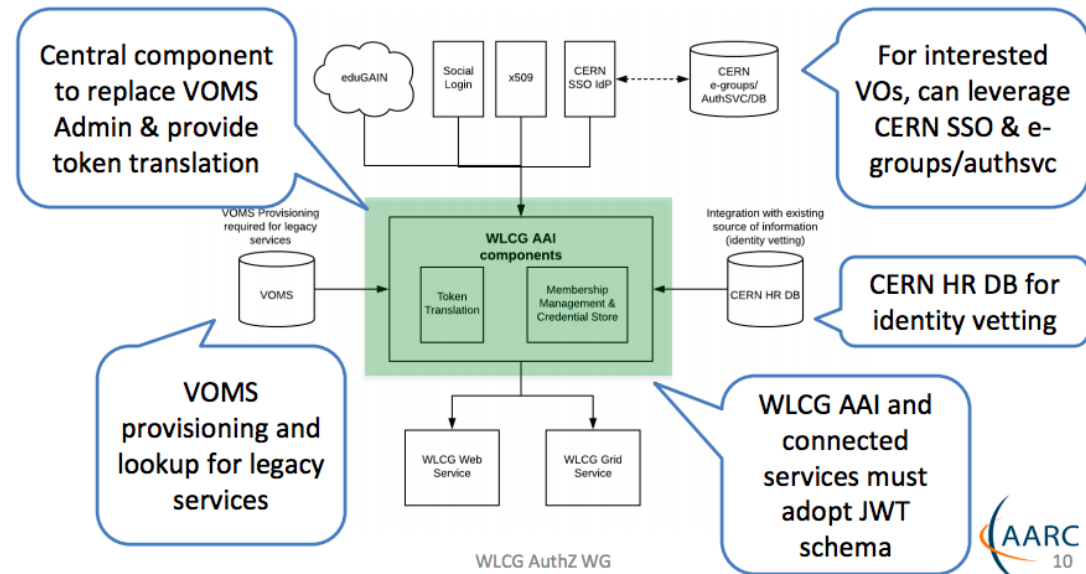
# DOMA and AAI

- AAI evolution in WLCG is driven by the WLCG AuthZ WG
  - Prototyping an architecture of which DOMA activities are one aspect
  - X509 free, based on Jason Web Tokens

- The WG collected requirements and is evaluating existing solutions for the WLCG MB

- All inline with the DOMA needs and strategy

- <http://cern.ch/go/9fXq>





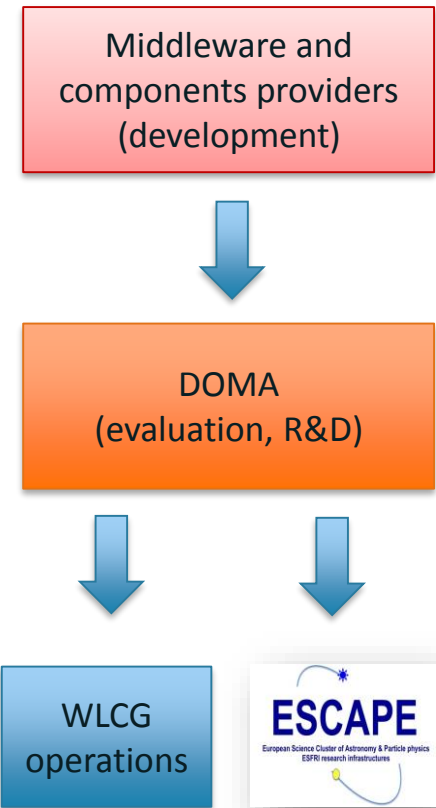
# Conclusions and future work

DOMA works bottom-up: does not define an architecture at this stage. We have indications of how facilities and services will look like and we define an R&D program to evaluate various technologies.

ESCAPE (WP2) will build a data lake prototype based on those technologies for HL-LHC and many other sciences

Rucio being used by ATLAS and CMS (with interest of other communities) opens a big opportunity and will play a central role in the data lake architecture

We will need a discussion about pledging resources as the current model has limitations



We plan to commission adiabatically many of those features much earlier than HL-LHC