

# Modernizing Third-Party-Copy

---

## Transfers in WLCG

A. Forti

On behalf of the DOMA TPC WG

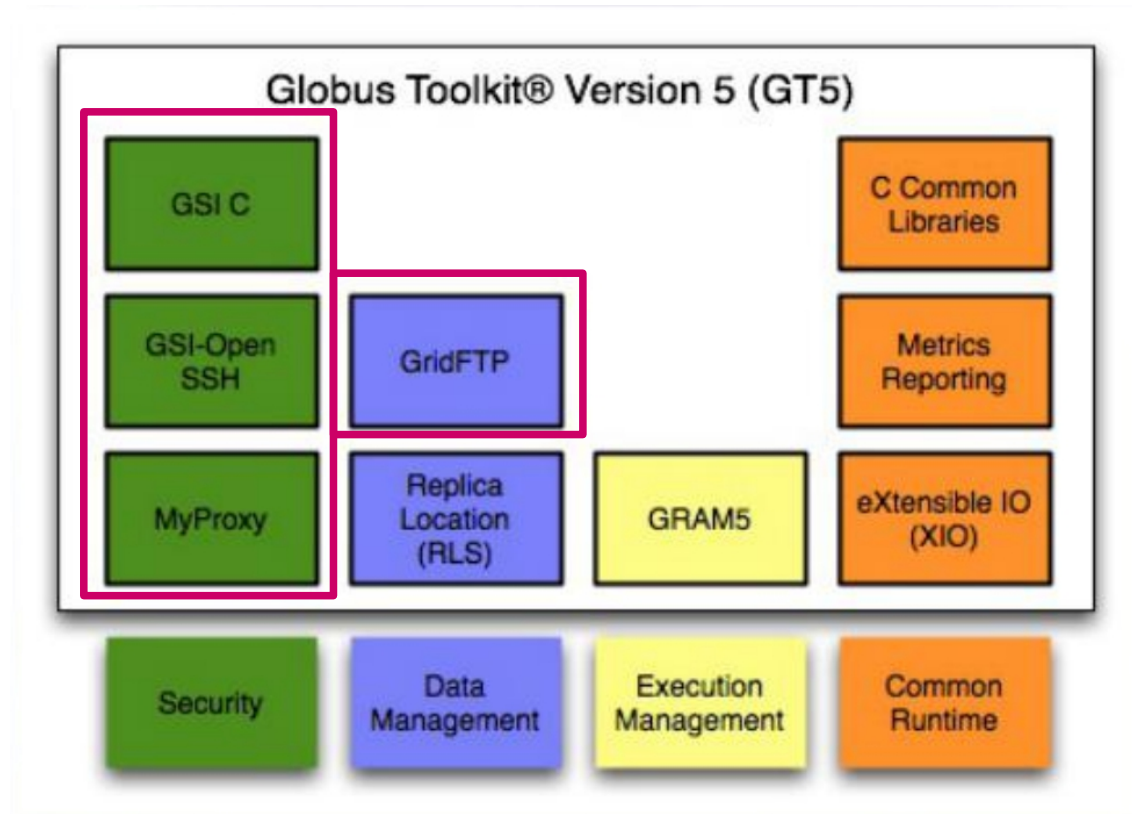
CHEP2019

05 November 2019



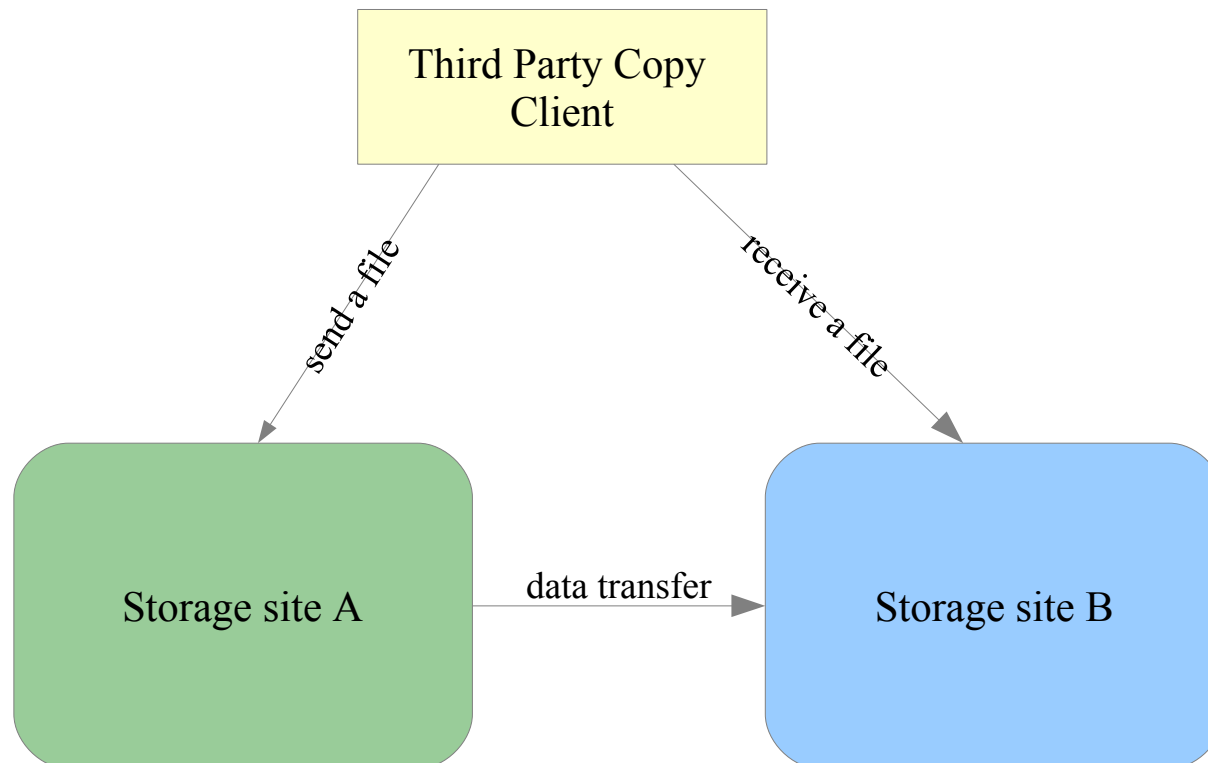
# Problem

- Grid was built around globus toolkit
- January 2018, the Globus team stopped supporting the open source Globus Toolkit included GridFTP
  - End of 2018 maintenance and security patches stopped too
- WLCG still uses
  - GridFTP, GSI, MyProxy
- Several services and clients affected
  - not only storage



# Third Party Copy

- WLCG experiment data transfers built around GridFTP and GSI
  - Bulk transfers done with Third Party Copy (TPC)
    - Request to move data from site A to site B



# DOMA TPC WG

- WLCG DOMA subgroup for TPC was decided at CHEP 2018.
- Goal is to have a working alternative to GridFTP at all sites.
  - Candidate protocols: xrootd and http
- Several services and clients involved
  - DPM, StoRM, dCache, XrootD, EOS, ECHO, Dynafed
  - Rucio, FTS
  - Gfal2, fts-client, storm-srm
- Required participation of a lot of people
  - Experiments, sites, storage, data management, clients developers



# Stages

- Phase 1: Ensure all storage implementations have a valid alternative in production
- Phase 2: Ensure rollout of non gridftp protocol at all sites with  $>3\text{PB}$
- Phase3: Rollout to remaining WLCG sites.



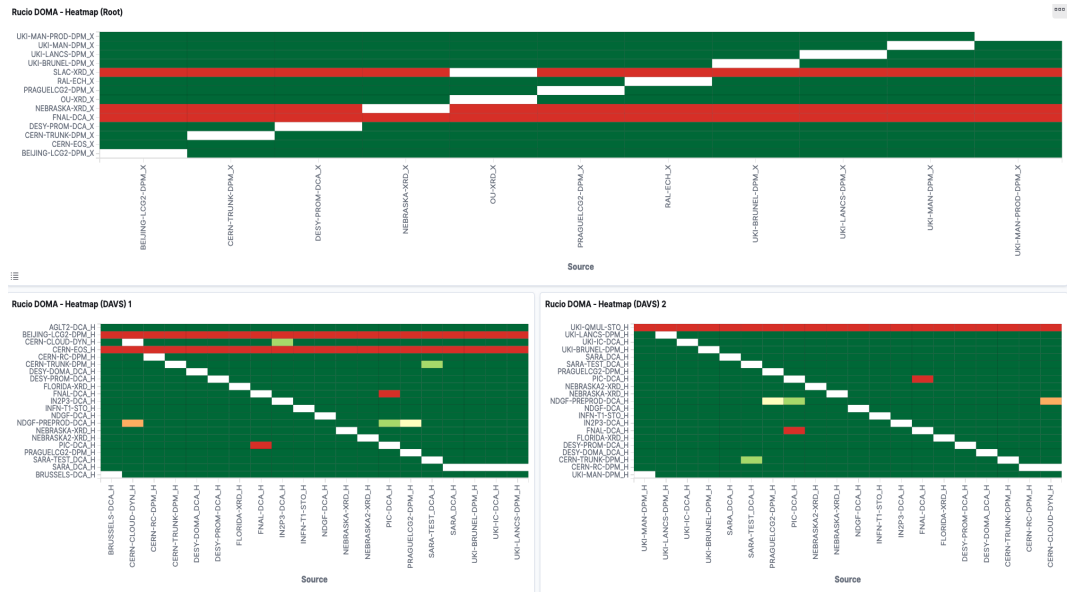
# Development

- Authorization
  - Xrootd: delegation
    - user proxy rather than servers or robot proxies
  - Http: tokens
    - In place of x509 proxy
  - Both simplify sites configuration and deployment
    - Not necessary to install public keys
- Checksums
  - Some storage implementations missed it
- Compatibility between storages
  - Different implementations talk to each other
    - Common http headers, dcache/xrootd delegation
  - Clients and servers have necessary features



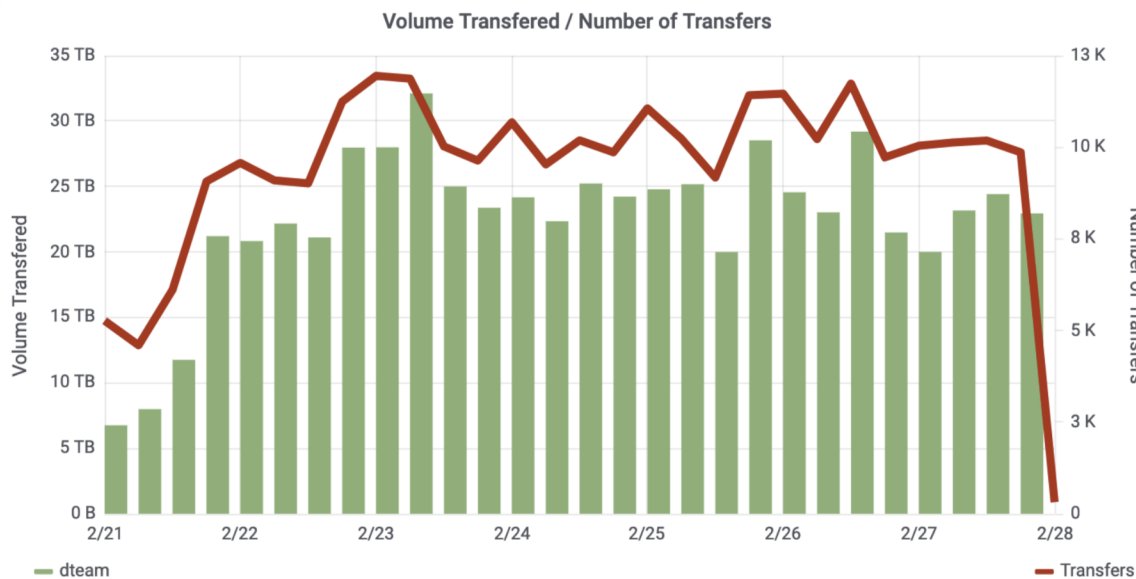
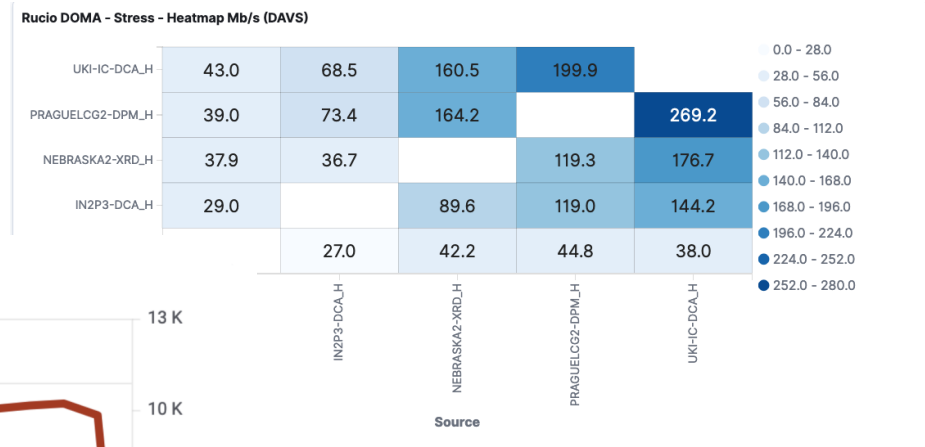
# Testing infrastructure

- Rucio instance + FTS dev specifically for DOMA TPC tests
  - 36 sites (testbeds, prod or both)
  - All sites participating are added with the supported protocols
- Functional: a small file every few hours for each pair of storages for each protocol
  - Any type of storage allowed
    - Prod and testbeds
- Kibana monitor
  - Heatmap for each protocol
  - Http had to be split in two
    - Too many sites



# Stress tests

- Still done with rucio but on a smaller number of sites
  - Stable storages with up-to-date storage middleware
  - 1 TB transferred repeatedly between each pair of sites to simulate larger transfers



Each bar is a 6-hour time window.



# Smoke tests

- Simpler scripts to run atomic tests
  - Results are not polluted by fall back solutions
    - push/pull, tpc/streaming
  - Capturing more diagnostic information
    - scripts runs several different tests to probe for different failure modes
  - Flexibility to test specific end points when it is needed
- We run such tests for each protocol every night and get a detailed summary report

-----PROBLEMATIC ENDPOINTS-----

SCORE	ENDPT	TYPE	UP	SRC	DST	DN	
16	CERN-EOS	EOS	F	-	F	F	0/4
16	SLAC	XrootD	F	-	F	F	0/4
4	RAL-LCG2	Echo	F	-	F	F	0/4
0	CALTECH	HDFS	P	P	F	P	3/4
0	BNL	dCache	F	-	F	F	0/4

-----ERROR DETAILS-----

# Where are we?

- Phase3: large scale deployment
  - All storages need to be upgraded to the latest version
- Major upgrade campaign
  - DPM and dCache have dedicated WLCG ops task forces to push the upgrade and help with new configuration
- FTS need to be upgraded too
  - FTS/gfal2/xrootd and davix and new library to support tokens



# Experiments activities

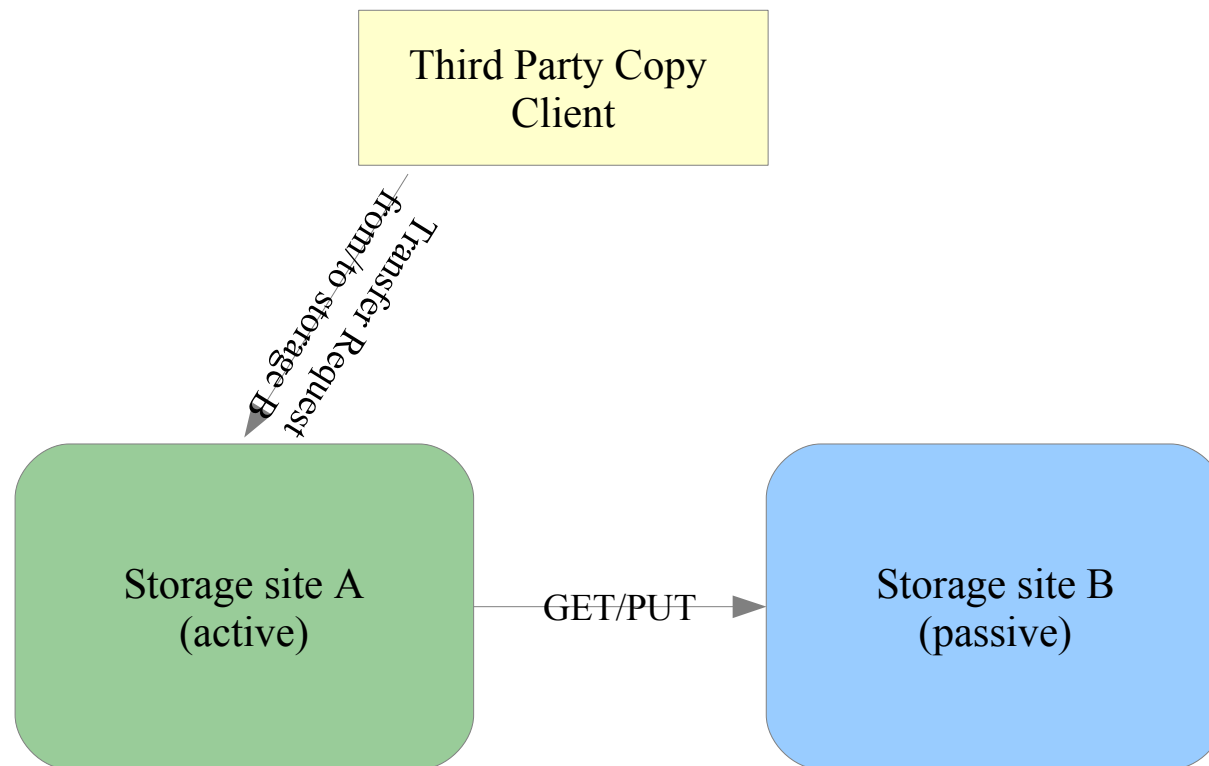
- USCMS has a target of 30% of production traffic done with a non-GridFTP protocol
  - at one site (Fermilab) by February 2020 using phedex
- ATLAS enabled TPC functional tests on its production infrastructure
  - TPC sites + T1s
    - requires upgrade of storage + FTS
    - http streaming hides tpc results

	AGLT2_SCRATCHDISK	AUSTRALIA-ATLAS_SCRATCHDISK	BNL-OSG2_SCRATCHDISK	CA-VICTORIA-WESTGRID-T2_SCRATCHDISK	CERN-PROD_SCRATCHDISK	DESY-HH_SCRATCHDISK	UNI-BONN_LOCALGROUPDISK
AGLT2_SCRATCHDISK	-	100%	-	100%	100%	100%	100%
AUSTRALIA-ATLAS_SCRATCHDISK	0%	-	0%	0%	0%	0%	0%
BNL-OSG2_SCRATCHDISK	-	100%	-	100%	100%	75%	100%
CA-VICTORIA-WESTGRID-T2_SCRATCHDISK	100%	100%	100%	-	100%	0%	100%
CERN-PROD_SCRATCHDISK	0%	0%	0%	0%	-	0%	0%
DESY-HH_SCRATCHDISK	0%	0%	0%	0%	0%	-	0%
PRAGUECG2_SCRATCHDISK	0%	0%	0%	0%	0%	0%	-
UKI-NORTHGRID-MAN-HEP_SCRATCHDISK	100%	100%	100%	100%	83%	83%	100%
UNI-BONN_LOCALGROUPDISK	100%	67%	0%	0%	0%	50%	100%



# GSI → Token

- Http protocol authorization developed with bearer token
  - Simplest form with storage generated macaroons
  - Work connected to WLCG Authz WG
  - Feasibility



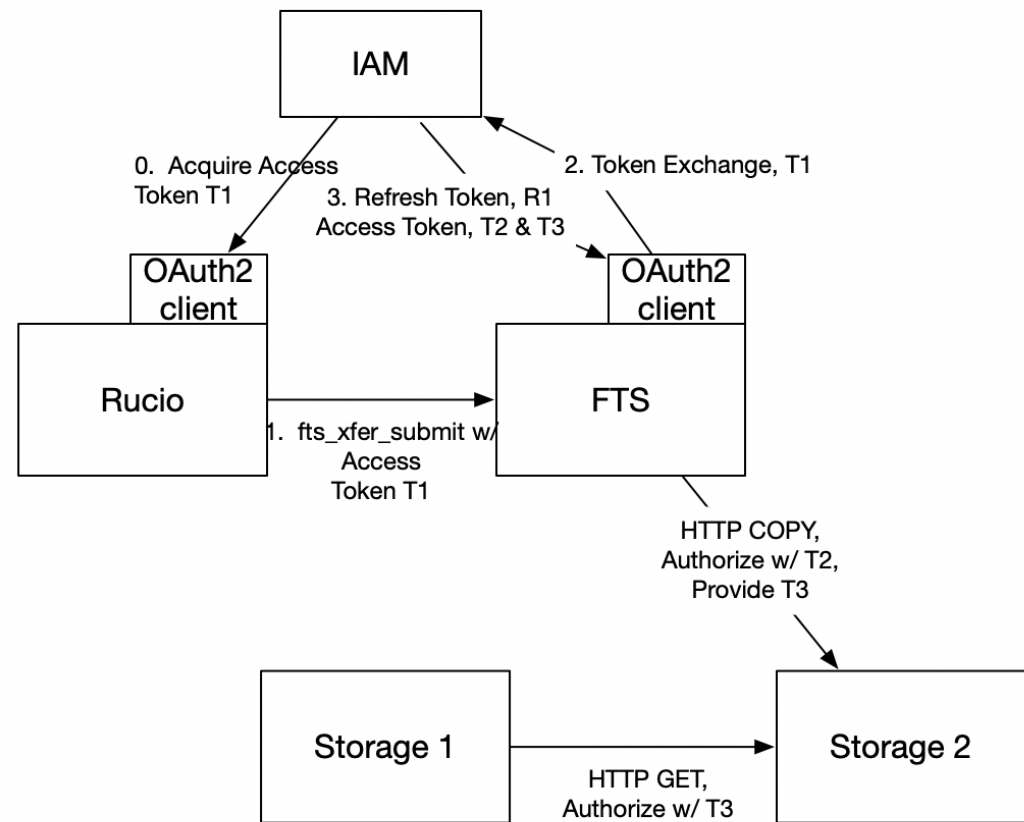
# Stages

- Phase 1: Ensure all storage implementations have a valid alternative in production
- Phase 2: Ensure rollout of non gridftp protocol at all sites with >3PB
- Phase3: Rollout to remaining WLCG sites.
- Phase4: Token based authz testbed **NEW!**



# Phase 4

- Token authorization testbed
- Since http can now handle tokens we added also setting up a testbed for the full token authorization chain
  - User → IAM → rucio → FTS → storage 1 → storage 2



B. Bockelmann

# Conclusions

- Added missing essential features
  - Checksums where missing
- Added new authentication methods
  - Delegation for xrootd
  - Tokens for http
- Added functionality to gfal/FTS/... clients
- Setup testing and monitoring infrastructure
  - Dedicated FTS and rucio instances
  - Involve sites, plug them in and debug problems
- Upgrading sites
- Working on token authz testbed



# Related talks

- B. Bockelmann  
Third-party transfers in WLCG using HTTP
- W. Yang  
Xrootd Third Party Copy for the WLCG and HL-LHC
- A. Ceccanti  
WLCG Authorisation; from X.509 to Tokens
- A. Ceccanti  
Beyond X.509: token-based authentication and authorization in

