

Bulk Data Transfer WG Update

29 June 2022



MORGRIDGE
INSTITUTE FOR RESEARCH
RESEARCH COMPUTING

FEARLESS SCIENCE

Bulk Data Transfer Working Group

The Bulk Data Transfer Working Group aims to coordinate, develop, and help integrate new technologies for the movement of (bulk) data between sites. In the last few months since the formation, the activities have focused on:

1. Coordinate with the storage providers and FTS team to test and deploy the REST-based API proposed for archive management. **Update today.**
2. Transition of bulk data transfers to token-based authorization. **Update today.**
3. Engage with networking projects / working groups to coordinate testing & integration of packet marking technologies.

Today's updates does not try to be exhaustive of all our activities (which still includes things like HTTP-TPC refinements!) but rather covers things I think are useful for the wider group.

The BDT team meets biweekly and all are welcome to join! [Last meeting's Indico...](#)

REST API for Archival Data

The REST API for archive (“tape”) management aims to provide a modern API for interacting with tape services.

- At the minimum, it replaces functionality provided by SRM for tape management without the baggage of SRM (e.g., custom non-TLS-based transport protocol).
- Not a 1-to-1 mapping of SRM’s XML to JSON; rather, many simplifications to stick to what we need.
- Ideas evolved from a dCache REST API; final product is not the same, however.
- [See current spec.](#)

BDT is starting to coordinate a set of [testbed endpoints](#).

Question: What’s a realistic timeline for switching off SRM?

CERN REST API planning (EOS, CTA and FTS)

(Slide update Cedric Caffy)

Tape REST API status

- **Tape REST API on EOS**
 - Tape REST API V0.1 released
 - Tape REST API V1 ready, to be released mid-July
 - Configurable options to enable Tape REST API interface
- **Tape REST API on FTS/GFAL**
 - Full REST support on the latest version of GFAL 2.21.0 (required from FTS)
 - FTS v3.12.0 to be released end-July
 - Support for “*EVICTION*” (*release*) and “*ARCHIVE MONITORING*” (*archiveinfo*)
 - FTS v3.12.1 to be released end-September
 - Support for FULL Tape REST API
 - FTS team has already FTS integration tests developed and running
 - Possibility to extend them to other endpoints (when available) to make sure FTS is compatible
- **CERN CTA Deployment plans**
 - “*EVICTION*” and “*ARCHIVE MONITORING*” is expected to be in production end-July 2022
 - FULL Tape REST API is expected to be in production in December 2022
- **RAL CTA Deployment plans (discussed at CERN with the CTA team)**
 - FULL Tape REST API is expected to be in production in February 2023

Discussion topic – Timeline?

- FTS support is expected late summer 2022.
- [dCache support](#) in golden release is scheduled for 8.2 -> August 2022?
 - Possibly some production endpoints during winter shutdown.
- CTA-based sites will be ready during winter shutdown 2022/2023.
- StoRM -> ??? Part of larger overhaul of the codebase, timeline unclear.
- The FTS and CTA teams will synchronize with Rucio in order to handle the subtle differences between the XRootD and HTTP tape protocols.
- Enable LHCb to replace GridFTP and SRM with HTTP in production by February 2023

Crystal ball predictions:

- Second half of 2022 will be for testbed instances and software verification.
- Leading-edge T1s will have production instances in 2023 available for tests.
- Site-by-site migration during 2023; ask all sites to finish by winter shutdown 2023/2024?
 - Is there need for more aggressive timelines?

Token Transition for BDT

Roughly, there are three steps for the token transition:

1. Deploy / verify support across storage technologies. <~ **You are here.**
2. Ensure FTS can accept/manage tokens and interact with storage endpoints as appropriate.
3. Update Rucio to generate the right tokens and orchestrate transfers through FTS.
 - To some extent, FTS is a “dumb” client and just executes the instructions provided by Rucio.
 - Additional step is to have Rucio manage tokens for non-third-party-copy transfers.
 - No updates for Rucio today – start of technical work is waiting on availability of personnel.

Note these are not really ‘steps’ but iterative, back-and-forth. E.g., FTS testing will reveal storage bugs; Rucio design work will potentially generate new requirements for FTS.

Storage Support for Tokens

BDT working group runs nightly tests against a token profile compliance suite.

- Suite itself has been improved and adapted to better cover our use cases in the last few months.
- Has uncovered several bugs in implementations.

Version availability:

- dCache 7.2 can be used immediately for all our use cases; 8.2 has some quality-of-life improvements sites may enjoy.
- XRootD has some significant fixes already committed; slated for the 5.5 release.
- EOS support is on par with XRootD 5 release.

BDT is working to generate suggested configurations for sites to follow.

FTS Support for Tokens

FTS has supported “tokens” for several years already.

- What’s being added now is better support for WLCG profile and toward the flows we think will be needed by Rucio.
 - Example of added complexity: likely will need two credentials per transfer, one for source site one for destination site.
- Once we have better token compliance for storage services, we’ll start to include FTS in our nightly testing.
- FTS will continue to adapt to changes and discoveries in the use of tokens.
 - All this should feed toward the DC2023!





MORGRIDGE
INSTITUTE FOR RESEARCH
RESEARCH COMPUTING

morgridge.org

FEARLESS SCIENCE