

# Minutes of the HTTP REST TAPE API meeting (3rd of December 2021)

## Useful link

- FTS presentation about FILEINFO being asynchronous vs synchronous:  
[https://indico.cern.ch/event/1100651/attachments/2358835/4026055/FTS\\_Tape\\_API\\_Fileinfo\\_Walkthrough.pdf](https://indico.cern.ch/event/1100651/attachments/2358835/4026055/FTS_Tape_API_Fileinfo_Walkthrough.pdf)

## Goal of the meeting

- Discuss about the fact that FILEINFO is synchronous or not
- Discuss about the GET /api/v1/stage/{id} result

## Discussions

Mihai (FTS) did a presentation to present his questions/concerns he had about FILEINFO

You can find the link of the presentation in the [Useful link](#) part of this document.

## FILEINFO will always be synchronous

After the presentation and some discussions, we agreed that the FILEINFO functionality of our API will always be synchronous.

A server must not refuse a request containing too many paths. It will only return the files from which it could get the information within a specific amount of time (defined by the server).

For example, if a client requests information about 1000 files and the server could only fetch the information for 500 during 1 second, the server will return the information about these 500 files. The client could then send another request about the files that were not returned by the server.

The synchronousResponse attribute of the JSON message submission will therefore be removed. The specification of the API will have to indicate that FILEINFO is synchronous as opposed to the STAGE and UNPIN functionalities that are asynchronous.

## Get the status of a subset of files that belong to a stage bulk-request

Mihai asked whether it would be possible to only get the status of a subset of files belonging to an already submitted stage request.

His suggestion was to use a POST `/api/v1/stage/{id}` with a JSON containing the list of paths of the files to get the status from.

As POST is not supposed to be used to get information about the resources, this poses a problem.

After some discussion, we all decided that for the first version of our API, a GET `/api/v1/stage/{id}` will always return the state of **ALL** the files that belong to the stage bulk-request.

## Future version of the API

### Between two stage progress polling, only send information about the files that were successfully staged

In a future version of the API, we can probably ask the client to provide a flag allowing us to know whether it is stateful or not. A stateful client will keep track of the files that are already staged and will therefore not be interested in getting information about them anymore. The server will then only send information about the files that were successfully staged.

# Submission of a STAGE bulk-request

For each file that belongs to the STAGE bulk-request, its path and additional information need to be passed

(<https://indico.cern.ch/event/1089473/attachments/2334949/3979754/WLCG%20REST%20APE%20API%20-%20STAGE%20bulk%20request%20submission.pdf>).

Some of these additional information might be part of the common schema, some might not and should be part of the `targeted_metadata` object. It was not decided which information will be part of the common schema and which will not be part of it.

## Targeted\_metadata object structure

Each metadata object within the `targeted_metadata` field will be identified by the REST API endpoint identifier to whom the metadata are targeted. This object will then contain the actual metadata.

For example:

```
"targeted_metadata":
{
  "eosctaatlas.cern.ch:8080": {
    "priority": 1
  },
  "eosctalhcb.cern.ch:8080": {
    "metadata2": "hello_world"
  }
}
```

The format of the endpoint identifier has not been validated yet.

The JSON object that has to be passed to the server by the client in order to submit a STAGE bulk-request would look like:

```
{
  "files": [
    {
      "path": "/eos/na62/test/test.txt",
      [...]
      "targeted_metadata":
      {
        "eosctaatlas.cern.ch:8080": {
          "priority": 1
        },
        "eosctalhcb.cern.ch:8080": {
          "metadata2": "hello_world"
        }
      }
    }
  ]
}
```

```
]
}
```

## Next meeting

Date: Friday the 3rd of December 3pm CET 8am CST

### To be discussed

CTA needs the following fields to stage files from tape:

- activity (string)
- priority (unsigned int)
- QoS (string)
- bringonlinedeadline (epoch timestamp)
- parentdataset (string given by Rucio)
- datasetsize (unsigned long given by Rucio)
- enduser (behind Rucio authentication)
- FTS job ID (UUID)
- FTS file ID (unsigned long)
- FTS endpoint (string)

What are the fields that we want to have in common (part as a first class citizen per-file attribute) and what are the fields we do not want to have in common?

What is the format we would like to have for the unique REST API endpoint identifier?

Continue the discussion on the other messages.

## Idea for a future version of the API

Advertise metadata audiences/targets in the .well-known mechanism so that FTS can filter before sending metadata to the experiments.