

# GERENCIAR PACOTES BAGIT

## CRIAR PACOTES BAGITS POR DATASETS

```
sudo su - glassfish
curl -H X-Dataverse-key:$API_TOKEN
http://localhost:8080/api/datasets/:persistentId/?persistentId=doi:10.48331/SCIELODATA.RJ1
HZ0
export ID=19300
export API_TOKEN='YOUR API TOKEN'
curl -X POST -H "X-Dataverse-key: $API_TOKEN"
http://localhost:8080/api/admin/submitDatasetVersionToArchive/$ID/1.0
```

## CRIAR PACOTES BAGITS DE TODOS OS DATASETS

Essa execução é feita uma vez por mês. Os pacotes bagits são gerados no diretório /preservacao/dataverse que é um ponto de montagem nfs em storage-scielo-232.scielo.org:/archivematica\_transfersource

```
export API_TOKEN='MEU TOKEN'
curl -X POST -H "X-Dataverse-key: $API_TOKEN"
'http://localhost:8080/api/admin/archiveAllUnarchivedDatasetVersions'
```

Uma vez que os pacotes são gerados precisamos organizar cada um dos datasets em pastas individuais. Para isso precisamos executar o script organizar-arquivos.sh que irá criar apenas as pastas

```
cd /preservacao/dataverse
sh -x organizar-arquivos.sh
```

Uma vez criadas as pastas, precisamos ajustar o nome da pasta, pois ele fica com a versão no final e o desejado é ter apenas o nome:

```
cd /preservacao/dataverse
sh -x normalizar.sh
```

Pode ser que o mês tenha muitos arquivos e que represente muito gigabyte como foi o caso de agosto de 2025 que chegou a 29Gb. Precisamos dividir elas em tamanho de 1GB. Foi criado um script chamado separador-pasta.sh. Antes de executar o script, crie a pasta que identifique o mês e o ano. Exemplo: scielodata-setembro-2025

```
mkdir /preservacao/dataverse/scielodata-setembro-2025
```

Agora edite o arquivo para ajustar a variável ORIGEM e A PREFIXO:

```
#!/bin/bash
# Diretório de origem
ORIGEM="/preservacao/dataverse/scielodata-setembro-2025"
# Prefixo das novas pastas
PREFIXO="scielodata-setembro-2025"

cd "$ORIGEM" || exit 1

i=1
total=0
destino=$(printf "%s-%02d" "$PREFIXO" "$i")
mkdir -p "$destino"

# Lista subpastas com tamanho em KB (para somar fácil)
du -sk * | sort -n | while read -r size name; do
    # Se passar de 1GB (1048576 KB), cria nova pasta
    if (( total + size > 1048576 )); then
        i=$((i+1))
        total=0
        destino=$(printf "%s-%02d" "$PREFIXO" "$i")
        mkdir -p "$destino"
    fi

    echo "Movendo $name para $destino (tamanho: $size KB)"
    mv "$name" "$destino/"
    total=$((total + size))
done
```

veja que neste exemplo as variáveis abaixo precisam ser ajustas para o mês que você precisa trabalhar:

```
ORIGEM="/preservacao/dataverse/scielodata-setembro-2025"  
PREFIXO="scielodata-setembro-2025"
```

Antes de executar o script mova as pastas para a pasta do mês:

```
mv doi-10-48331-scielodata-* scielodata-setembro-2025/  
./separador-pasta.sh
```

## TROUBLESHOOTING

### Version was already submitted for archiving.

Quando você envia para criar o pacote em uma data versão, não é possível criar o mesmo sem uma intervenção. O erro que dá é:

```
{"status":"ERROR","message":"Version was already submitted for archiving."}
```

Para enviar novamente siga os passos:

```
sudo su - postgres  
\c dvndb  
SELECT  
    id, versionnumber, minorversionnumber, versionstate,  
    archivetime, archivalcopylocation, externalstatuslabel, archivenote  
FROM datasetversion  
WHERE dataset_id = '13150'  
ORDER BY versionnumber DESC, minorversionnumber DESC;
```

Veja que o resultado foi:

```
dvndb=# ORDER BY versionnumber DESC, minorversionnumber DESC;  
 id | versionnumber | minorversionnumber | versionstate | archivetime | archivalcopylocation | ex  
alstatuslabel | archivenote  
-----+-----+-----+-----+-----+-----+-----  
2209 | 1 | 0 | RELEASED | | {"status":"failure","message":"Bag not transferred"} |  
(1 row)  
dvndb=#
```

Vamos agora atualizar o registro para limpar esta mensagem:

```
UPDATE datasetversion
SET archivetime = NULL,
    archivalcopylocation = NULL,
    externalstatuslabel = NULL,
    archivenote = NULL
WHERE id = 2209;
```

Resultado:

```
dvndb=# ORDER BY versionnumber DESC, minorversionnumber DESC;
 id | versionnumber | minorversionnumber | versionstate | archivetime | archivalcopylocation | externalstatuslabel | archivenote
-----+-----+-----+-----+-----+-----+-----+-----
 2209 |             1 |                   0 | RELEASED    |             |                     |                     |
(1 row)

dvndb=# █
```

Agora vamos executar novamente o comando:

```
sudo su -glassfish
export ID=13150
export API_TOKEN='minha api'
curl -X POST -H "X-Dataverse-key: $API_TOKEN"
http://localhost:8080/api/admin/submitDatasetVersionToArchive/$ID/1.0
```

Outra forma para limpar é:

```
export API_TOKEN=xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxx
export SERVER_URL=https://demo.dataverse.org
export PERSISTENT_IDENTIFIER=doi:10.5072/FK2/7U7YBV
export VERSION=1.0

curl -H "X-Dataverse-key: $API_TOKEN" -X DELETE
"$SERVER_URL/api/datasets/:persistentId/$VERSION/archivalStatus?persistentId=$PERSISTENT_IDENTIFIER"
```

Link: <https://guides.dataverse.org/en/latest/api/native-api.html#delete-the-archival-status-of-a-dataset-by-version>

**Erro {"status":"failure","message":"Bag not transferred"}**

Este erro aconteceu quando eu tentei criar os bagit e o usuário não tinha permissão para escrever no diretório /preservacao/dataverse. Uma vez que deu erro alguns pacotes ficaram com o status : {"status":"failure","message":"Bag not transferred"}

1 Primeiro: verifique antes de atualizar (modo seguro)

Antes de alterar qualquer coisa, rode este SELECT para confirmar exatamente **quais registros serão afetados**:

```
SELECT
    id, dataset_id, versionnumber, versionstate, archivetime, archivalcopylocation
FROM datasetversion
WHERE archivalcopylocation::text LIKE '%"Bag not transferred"%';
```

Isso lista **apenas as versões com erro de arquivamento**, garantindo que você não vai limpar dados corretos.

2 Atualização segura (modo em lote)

Depois de confirmar, execute:

```
UPDATE datasetversion
SET
    archivetime = NULL,
    archivalcopylocation = NULL,
    externalstatuslabel = NULL,
    archivenote = NULL
WHERE archivalcopylocation::text LIKE '%"Bag not transferred"%';
```

Esse comando:

- **limpa os campos de arquivamento** que ficaram marcados com erro,
- sem afetar versões bem-sucedidas ou DRAFTs normais.

Feito isso, faça novamente o processo de criação dos pacotes.