

Controlling Total File Storage Size in AIRNow Drupal

INTRODUCTION

This is a pair of procedures to control the total file size on the TOMEd AirNow Drupal server for issues AIR-423 & AIR-441. This is attempted by “Removing Unused Files” one a month, and “Migrating the 5 Largest Files” once a quarter.

If the files directory makes the TOMEd app too big to push to cloud.gov, then try “Migrating the 5 Largest Files” procedure first.

Big Picture:

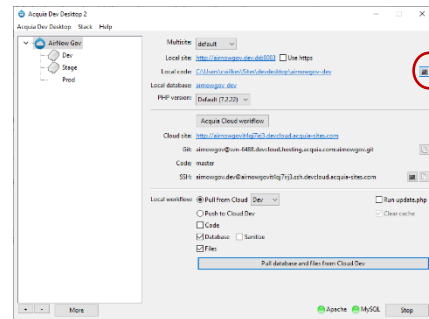
1. Once a month: Removing Unused Files
2. Once a quarter: Migrating the 5 Largest Files

PROCEDURE: REMOVING UNUSED FILES

MONTHLY

Set-up: Pull the files from DevCloud

1. **Make sure that no one else will be changing Content on DevCloud during this Procedure!**
2. Pull latest version of the site Content (database) *and* Files from Acquia DevCloud using DevDesktop while **off** the EPA VPN network.
3. Use DevDesktop to open a *drush* window on your local host. Click on the local host *drush* console button... Note: There are two *drush* console buttons; use the upper one.



Find the Files: drush sqlq

4. Learn how “drush sqlq” works. Find some File IDs (fid) and paths using *drush* SQL queries.

```
// Example: List of Unused files with a filename starting with 'wildfire'.
drush sqlq "SELECT fid, status, filename, uri FROM file_managed WHERE NOT EXISTS (SELECT * FROM file_usage WHERE file_managed.fid = file_usage.fid) AND filename LIKE 'wildfire%'"
```

```
// Details for ALL Unused files
drush sqlq "SELECT fid, status, filename, uri FROM file_managed WHERE NOT EXISTS (SELECT * FROM file_usage WHERE file_managed.fid = file_usage.fid) AND status = 1"
```

5. Now focus on the PDF files from Last Month ONLY by changing the date value in the last clause of this sql statement. (Maybe check a couple of months back while you're set-up...)

```
// PDF files from a specific month
drush sqlq "SELECT fid, status, uri, filesize FROM file_managed WHERE NOT EXISTS (SELECT * FROM file_usage WHERE file_managed.fid = file_usage.fid) AND status = 1 AND uri LIKE '%pdf' AND uri LIKE '%2021-12%'"
```

Delete the Files

6. PDF files in the list from the step above are NOT used in the Drupal Content database. Delete each one from the path shown in the output.
7. Delete them from the Content database using this SQL statement while changing the fid value to the values from the “Last Month ONLY” output.

```
// Delete a single PDF file for the given fid
drush sqlq "DELETE FROM file_managed WHERE file_managed.fid = <fid>"
```

Controlling Total File Storage Size in AIRNow Drupal

8. Verify that you deleted them by re running the command the “Last Month ONLY” step. There should be no output.

Push the Content and Files back

9. Push the Content database **and** Files back to Acquia DevCloud using DevDesktop while **off** the EPA VPN network.

Done!

PROCEDURE: MIGRATING THE 5 LARGEST FILES

QUARTERLY

Which Files?

1. Open a PowerShell window and change to the current “files” directory for your local Dev Desktop:

```
cd C:\Users\cwilkes\Sites\devdesktop\airnowgov-dev\docroot\sites\default\files
```

2. Run this PowerShell command to retrieve the 5 largest file names & subdirectories:

```
Get-ChildItem -Path "*.pdf","*.pptx" -Recurse -File | Sort-Object -Property Length -Descending | Select-Object -first 5
```

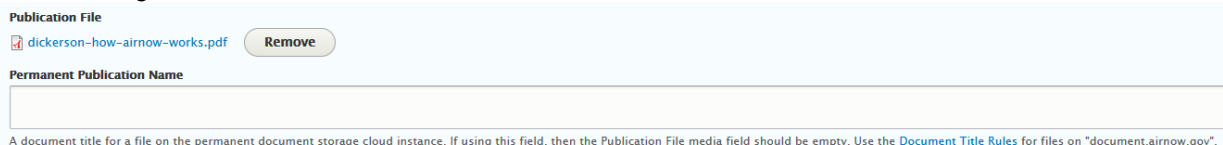
Put the 5 files on “document.airnow.gov”

3. Pull a fresh copy of the “document” repository found on BitBucket.
4. Copy the 5 files above to your local .git repository for “document”. Convert any PowerPoint presentations to PDF.
5. Verify that each of the 5 files are NOT already in the “document” repository. If they are there, then note the Permanent File Name for use below.
6. Rename each new file using the rules in the “Document Naming Rules.txt” file that can be found in the repository. This is the file’s “Permanent File Name”.
7. Add, commit, and push to the “document” remote repository the 5 renamed documents.
8. Using Blue-Green Deployment, “cf push” the “document.airnow.gov” website on cloud.gov

Update the Publication on DevCloud

9. On DevCloud, Select Structure and then Taxonomy.
10. Click on the Publications “List terms” button at the end of the row.
11. Find the desired Publication Group Name and click “Edit”.
12. Scroll or search to find the publication being updated.

Before changes...



The screenshot shows a web form for editing a publication. At the top, it says "Publication File" and shows a file named "dickerson-how-airnow-works.pdf" with a "Remove" button next to it. Below that is a text input field labeled "Permanent Publication Name". At the bottom, there is a small note: "A document title for a file on the permanent document storage cloud instance. If using this field, then the Publication File media field should be empty. Use the [Document Title Rules](#) for files on "document.airnow.gov"."

13. Insert the “Permanent Publication Name” of the file that was stored on “document.airnow.gov”.
14. Click “Remove” to delete the file’s references in the Drupal database.

Controlling Total File Storage Size in AIRNow Drupal

After changes...

Publication File

No file selected.

One file only.

256 MB limit.

Allowed types: txt pdf pptx png jpg exe.

Permanent Publication Name

how-sacramento-brings-the-message-to-the-masses.pdf

A document title for a file on the permanent document storage cloud instance. If u

15. Update these values if empty or changing.
 - a. Publications Pages = the number of pages in the publication; NOT automatic
 - b. Publications Size = KB or MB values of the Publication size; NOT automatic
16. Scroll to the very bottom of the page and click "Save"
17. Go to the Publication Metadata page on DevCloud and verify that the link now goes directly to the correct PDF file on "document.airnow.gov".
18. REPEAT for all 5 files.
19. Test & verify.

Done!

DOCUMENT HISTORY

2022-01-20 Removing Unused Files copied from an existing procedure. Initial version by Chris Wilkes.
2022-01-25 Dropped "Fancy File Delete"; Now using drush sqlq commands by Chris Wilkes