How do I stop duplicate harvesting from CONTENTdm and DSpace through the Digital Collection Gateway?

Last updated: Fri, 12 Jul 2019 10:21:08 GMT

Answer

If you are harvesting CONTENTdm collections to WorldCat.org through the Digital Collection Gateway from two repositories you may see duplication in the records since the repository URL is the identifier for harvesting. If your records have already been harvested to WorldCat.org from DSpace prior to being migrated to CONTENTdm, they will be harvested again and the record will be duplicated.

In order to remove the duplication, you will need to do the following things:

1. Modify your Digital Collection Gateway schedule for one of the repositories and/or remove the site completely; and then,
2. Send this list of OCNs to the Bibchange team, notifying them that these are duplicates created by the Digital Collection Gateway and need to be deleted from WorldCat. If there are holdings on the record, Bibchange will not be able to delete them until all holdings are removed.
3. Once they've notified you that the duplicates have been purged from WorldCat, resync your collection.
4. If duplicates continue to populate, you will need to have Bibchange clear them out again, then delete the collection and resync it from scratch.

Additional information

While Bibchange is deleting the duplicates, ensure that you do not have a sync scheduled to run. If you have a sync scheduled, either cancel it or send the duplicates to Bibchange after the sync process has run.