Pulp - Refactor #8481

Revisit how published metadata are stored

03/30/2021 09:48 PM - dalley

Status:	CLOSED - NOTABUG		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:	dkliban@redhat.com		**% Done:**	0%
Category:			**Estimated time:**	0:00 hour
Sprint/Milestone:				
Platform Release:			**Tags:**	Wishlist
Groomed:	No		**Sprint:**	Sprint 96
Sprint Candidate:	No		**Quarter:**	

Description

Published Metadata inherits from Content, but are never assigned to repositories - essentially we just ignore them during orphan cleanup such that they are "leaked" forever.

This has led to workarounds needed in several places such as the universal content list in addition to orphan cleanup.

It seems that this would also mean that published metadata files are leaked forever even after their publications are destroyed.

That is... not great. We should investigate ways to improve this situation.

History

#1 - 05/01/2021 04:17 PM - dalley

- Description updated

#2 - 05/12/2021 03:25 PM - dkliban@redhat.com

- Status changed from NEW to ASSIGNED

- Assignee set to dkliban@redhat.com

#3 - 05/12/2021 03:26 PM - dkliban@redhat.com

- Sprint set to Sprint 96

#4 - 05/12/2021 10:17 PM - dkliban@redhat.com

- Status changed from ASSIGNED to CLOSED - NOTABUG

This doesn't seem to be a problem with how PublishedMetadata is stored.

Orphan cleanup task removes all content that is not part of any repository version[0]. It then cleans up any artifact that is not part of content[1].

PuplishedMetadata content is removed when associated publication is removed[2]. Any artifact associated with PublishedMetadata will then be deleted by the orphan cleanup task.

I tested this with the file plugin and the PULP_MANIFEST file is removed by orphan cleanup task.

```
$ pulp file remote create --name repo2 --url https://fixtures.pulpproject.org/file/PULP_MANIFEST
$ pulp file repository create --name repo2 --remote repo2 --autopublish
$ pulp file repository sync --name repo2
$ tree /var/lib/pulp/media/artifact/
/var/lib/pulp/media/artifact/
├── 38
│   └── ada38d05e98d6f7042311ad8af03946db148b6943a978ea34964bebdf44419
├── 3c
│   └── dce35167043d3838ef8fed64606fd5917fccaaa2be15c15f413683d03692fb
├── 4e
├── c0
│   └── 17ed9767546fe0edd174853391b763da5cc4eb911409dec49423a9c3a592a6
└── dc
    └── 9fe048a0ee80ea9ef3f574412ce35aa746542fbe75ca0ca9753e45a127dfcf
```

```
5 directories, 4 files


$ pulp file repository destroy --name repo2
Started background task /pulp/api/v3/tasks/8bd6fbe9-8bc3-4fa8-818d-67c889195fd7/
.Done.
$ pulp orphans delete
Started background task /pulp/api/v3/tasks/8f257451-ff7b-44df-a3df-ea72dc3de1d5/
.Done.
{
  "pulp_href": "/pulp/api/v3/tasks/8f257451-ff7b-44df-a3df-ea72dc3de1d5/",
  "pulp_created": "2021-05-12T20:32:47.949956Z",
  "state": "completed",
  "name": "pulpcore.app.tasks.orphan.orphan_cleanup",
  "logging_cid": "8401ae9e6c54433685c69185b4618643",
  "started_at": "2021-05-12T20:32:48.019892Z",
  "finished_at": "2021-05-12T20:32:48.190797Z",
  "error": null,
  "worker": "/pulp/api/v3/workers/6717762c-20d4-4307-a158-22d6d1b12238/",
  "parent_task": null,
  "child_tasks": [],
  "task_group": null,
  "progress_reports": [
    {
      "message": "Clean up orphan Content",
      "code": "clean-up.content",
      "state": "completed",
      "total": 3,
      "done": 3,
      "suffix": null
    },
    {
      "message": "Clean up orphan Artifacts",
      "code": "clean-up.content",
      "state": "completed",
      "total": 4,
      "done": 4,
      "suffix": null
    }
  ],
  "created_resources": [],
  "reserved_resources_record": []
}
$ tree /var/lib/pulp/media/artifact/
/var/lib/pulp/media/artifact/
├── 38
├── 3c
├── 4e
├── c0
└── dc

5 directories, 0 files
```

[0] https://github.com/pulp/pulpcore/blob/master/pulpcore/app/tasks/orphan.py#L48-L61

[1] https://github.com/pulp/pulpcore/blob/master/pulpcore/app/tasks/orphan.py#L66-L91

[2] https://github.com/pulp/pulpcore/blob/master/pulpcore/app/models/publication.py#L211

#5 - 05/13/2021 11:25 AM - dkliban@redhat.com

I performed a similar test with the RPM plugin and did not see any PublishedMetadata artifacts remaining after running the orphan delete task.

```
$ pulp rpm remote create --name repo3 --url https://fixtures.pulpproject.org/rpm-signed/
$ pulp rpm repository create --name repo3 --remote repo3 --autopublish
$ pulp rpm repository sync --name repo3
$ pulp orphans delete
```