

RPM Support - Issue #9328

repository metadata not getting published for some repos which have ".." in location_href (points outside of repo)

09/01/2021 05:11 PM - mwaqas@bloombergindustry.com

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	High	Due date:	
Assignee:	dalley	Estimated time:	0:00 hour
Category:		Groomed:	No
Sprint/Milestone:	3.16.0	Sprint Candidate:	No
Severity:	3. High	Tags:	
Version:		Sprint:	Sprint 105
Platform Release:		Quarter:	
OS:			
Triaged:	Yes		
Description			
I created multiple repos and selected default setting but metadata does not get published. Packages directory is missing from published repo. 'manually republishing metadata fixes it but I am afraid it will become stale.			
Related issues:			
Copied to RPM Support - Backport #9392: Backport #9328 "repository metadata n...		CLOSED - CURRENTRELEASE	

Associated revisions

Revision 59d67f8a - 09/13/2021 09:48 PM - dalley

Dealt with another situation incompatible with mirror syncing

Some repos use location_href values that point outside of the repository. We cannot support this in mirror mode because we cannot write paths outside of the repository namespace.

closes: #9328 <https://pulp.plan.io/issues/9328>

History

#1 - 09/02/2021 05:08 AM - dalley

- Project changed from Pulp to RPM Support
- Priority changed from Normal to High
- Sprint set to Sprint 104

#2 - 09/02/2021 05:38 AM - dalley

@mwaqas@bloombergindustry.com Which versions of pulpcore and pulp_rpm are you using?

#3 - 09/02/2021 05:51 AM - dalley

- Status changed from NEW to ASSIGNED
- Assignee set to dalley

Confirmed w/ master branch

#4 - 09/02/2021 05:09 PM - dalley

- Triaged changed from No to Yes

#5 - 09/03/2021 08:47 PM - ttereshc

- Subject changed from repository metadata not getting published to repository metadata not getting published for some repos sync from a mirror list

#6 - 09/03/2021 08:47 PM - ttereshc

- Subject changed from repository metadata not getting published for some repos sync from a mirror list to repository metadata not getting published

for some repos synced from a mirror list

#8 - 09/05/2021 09:38 PM - mwaqas@bloombergindustry.com

dalley wrote:

@mwaqas@bloombergindustry.com Which versions of pulpcore and pulp_rpm are you using?

```
rpm -qa |grep -i pulpcore
tfm-rubygem-pulpcore_client-3.14.1-1.el7.noarch
pulpcore-selinux-1.2.4-1.el7.x86_64
python3-pulpcore-3.14.4-1.el7.noarch
rpm -qa |grep -i pulp_rpm
tfm-rubygem-pulp_rpm_client-3.13.3-1.el7.noarch
```

#9 - 09/10/2021 12:28 AM - rchan

- Sprint changed from Sprint 104 to Sprint 105

#10 - 09/10/2021 10:54 PM - dalley

Also reproduced with https://packages.rundeck.com/pagerduty/rundeck/rpm_any/rpm_any/x86_64

#11 - 09/10/2021 11:28 PM - dalley

- Subject changed from repository metadata not getting published for some repos synced from a mirror list to repository metadata not getting published for some repos synced from behind Amazon CDN

https://packages.rundeck.com/pagerduty/rundeck/rpm_any/rpm_any/x86_64 is not a mirrorlist repo, so that is not the issue.

But both of these repos are hosted behind the Amazon Cloudfront CDN, which is doing some really crazy and annoying obfuscation (presumably for purposes of DDoS prevention). I don't have proof yet but I suspect that's part of the problem.

edit: I was wrong, that's not the reason. The reason is that these repos use location_href values that point backwards outside of the actual repo. Pulp cannot mirror these paths, so the only solution is to throw an error and tell the user to disable mirroring, so that the location_href values in the metadata can be rewritten inside the path namespace of the distribution.

#12 - 09/11/2021 12:01 AM - dalley

The solution is to disable "mirror-on-sync". And the "bugfix" here should be to add an hard-error that fails for repositories like this one, and tells the user to disable mirroring.

The "mirror" option makes a complete copy of the metadata and retains the same package layout as the original repo. But because the original layout keeps all of the package outside of the repo itself, mirror can't possibly work. It would be putting the files outside of the distribution base_path and polluting / interfering with the global namespace, and we cannot allow that.

#13 - 09/11/2021 03:43 AM - dalley

- Status changed from ASSIGNED to POST

#14 - 09/11/2021 04:15 AM - pulpbot

PR: https://github.com/pulp/pulp_rpm/pull/2121

#15 - 09/11/2021 04:27 AM - dalley

- Subject changed from repository metadata not getting published for some repos synced from behind Amazon CDN to repository metadata not getting published for some repos which have ".." in location_href (points outside of repo)

#16 - 09/11/2021 07:11 AM - dalley

- Sprint/Milestone set to 3.14.4

#17 - 09/13/2021 09:49 PM - dalley

- Status changed from POST to MODIFIED

Applied in changeset [59d67f8a598a98980a8487a5222842e67556dd2c](https://github.com/pulp/pulp_rpm/pull/2121).

#18 - 09/13/2021 09:50 PM - dalley

- Sprint/Milestone changed from 3.14.4 to 3.16.0

#19 - 09/13/2021 09:51 PM - dalley

- Copied to Backport #9392: Backport #9328 "repository metadata not getting published for some repos which have ".." in location_href (points outside of repo)" to 3.14.z added

#20 - 10/20/2021 09:07 PM - pulpbot

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE

Files

unpublished_metadata.png	47.6 KB	09/01/2021	mwaqas@bloombergindustry.com
repo_settings.png	355 KB	09/01/2021	mwaqas@bloombergindustry.com