

RPM Support - Issue #2048

Errata update failure during sync or upload

06/30/2016 07:30 PM - ttereshc

Status: CLOSED - CURRENTRELEASE	Start date:
Priority: Urgent	Due date:
Assignee: ttereshc	Estimated time: 0:00 hour
Category:	
Sprint/Milestone:	
Severity: 3. High	Groomed: No
Version: 2.8.5	Sprint Candidate: No
Platform Release: 2.8.6	Tags: Pulp 2
OS:	Sprint: Sprint 4
Triaged: Yes	Quarter:

Description

- Create repo where at least one erratum has no `updated` field or the `updated` field with the wrong datetime format.
- Sync it
- Do something to make the next sync operational (update importer, remove some unit)
- Sync it again and see the error

Task Failed

```
Could not parse errata `updated` field: expected format '%Y-%m-%d %H:%M:%S'.  
Fail to update the existing erratum SOME_ERRATUM_ID.
```

Errata is not updated if the `updated` field can not be parsed or absent.
This behavior was introduced with [this fix](#).

The malformed `updated` field or the absence of it is found in some RHEL and EPEL repositories/

Related issues:

Related to Packaging - Task #2083: Issues common to 2.9.1 and 2.8 stream	CLOSED - NOTABUG
Has duplicate Pulp - Issue #2070: Could not parse errata `updated` field: exp...	CLOSED - DUPLICATE

Associated revisions

Revision 3a1ebcec - 06/30/2016 11:11 PM - ttereshc

Make the parsing of the erratum updated field more tolerant

closes #2048 <https://pulp.plan.io/issues/2048>

Revision 637665cb - 07/08/2016 05:33 PM - bmbouter

PR review fixes for 2048

- Brings back RPM1007
- Causes update_needed() to return False in the case that either updated field is an empty string
- Adjust the logging to cause 1 line to be logged for all errata with non-empty and not recognized values for the updated field.

<https://pulp.plan.io/issues/2048> re #2048

History

#2 - 06/30/2016 07:50 PM - ttereshc

There are two options to fix it:

1. "can't parse, so never update"
 - that was the case for all errata before 2.8.5
 - we avoid a potential inconsistency between the erratum pkglist and the rest of the erratum metadata
 - but we do not give users an opportunity to have a recent erratum version if the `updated` field is wrong
2. "can't parse, so always update"
 - we give an opportunity to update the erratum if the `updated` field is wrong
 - there are potentially several scenarios when we can end up with wrong data in the erratum.

Bad scenario 1:

- create repo with some feed pointing to the repo with new errata version and bad `updated` field.
- sync it, now we have bad `updated` in the db.
- update repo with another feed which points to the repo with the same erratum but old version
- all the erratum metadata and package list will be overwritten

If we copied our repository before updating the feed, our update of the erratum in db will affect not only the repo we updated but also all the copied ones.

I think multiple copies of the repo is a common use case for our customers.

Bad scenario 2:

- create repo with some feed pointing to the repo with **new** errata version and bad `updated` field.
- sync it, now we have bad `updated` in the db.
- create another repo with some feed pointing to the repo with **old** errata version and bad `updated` field.
- sync it
- the erratum metadata will be overwritten, but package lists will be merged, so in db there will be old metadata and both old and new pkglist (the latter is correct)

#3 - 06/30/2016 07:51 PM - mhrivnak

- *Sprint/Milestone set to 22*
- *Triaged changed from No to Yes*

#4 - 06/30/2016 07:59 PM - bmbouter

Neither of these are great options, but I think I prefer option 1 because in terms of Pulp's consistency it's good to "avoid a potential inconsistency between the erratum pkglist and the rest of the erratum metadata".

If an erratum contains bad data, we should open an issue against the CDN tooling so that the data quality problem can be fixed.

Option 2 leaves Pulp in an inconsistent state in several cases so we probably shouldn't do that one.

#5 - 06/30/2016 08:01 PM - mhrivnak

I agree. Brian brings up a great point that we should encourage content creators to clean up the data and make the updated field parsable. That would allow their content to get updated if we go with option 1.

#7 - 06/30/2016 11:18 PM - ttereshc

- Status changed from ASSIGNED to POST

https://github.com/pulp/pulp_rpm/pull/925

#8 - 07/06/2016 02:47 PM - Anonymous

ttereshc wrote:

https://github.com/pulp/pulp_rpm/pull/925

This happened to me when trying to sync an EPEL 7.x repository on a brand new pulp server

#9 - 07/06/2016 06:52 PM - mhrivnak

- Priority changed from High to Urgent

Based on the user impact we've seen from pulp-list traffic, we decided to hold 2.8.6 another half-day to get this included.

#10 - 07/08/2016 06:23 PM - bmbouter

- Platform Release set to 2.8.6

A new PR is now being used to merge this. It has the original commit from ttereshc and another one from myself.

https://github.com/pulp/pulp_rpm/pull/928/files

#11 - 07/08/2016 07:42 PM - ttereshc

- Status changed from POST to MODIFIED

- % Done changed from 0 to 100

Applied in changeset [3a1ebcec5883d829ba5264dc24957331932ea685](#).

#12 - 07/08/2016 08:13 PM - bmbouter

- Has duplicate Issue #2070: Could not parse errata `updated` field: expected format '%Y-%m-%d %H:%M:%S'. added

#13 - 07/12/2016 10:01 PM - semyers

- Status changed from MODIFIED to 5

#14 - 07/15/2016 09:04 PM - pthomas@redhat.com

The Upgrade automation jobs have been passing.

#15 - 07/15/2016 10:09 PM - pthomas@redhat.com

- Status changed from 5 to 6

Verified

Synced repos with errata pre upgrade to 2.8.6.

Verified that the same repo can be re synced after upgrade to 2.8.6

#16 - 07/18/2016 09:21 PM - semyers

- Status changed from 6 to CLOSED - CURRENTRELEASE

#17 - 07/19/2016 09:00 PM - semyers

- Related to Task #2083: Issues common to 2.9.1 and 2.8 stream added

#18 - 03/08/2018 07:54 PM - bmbouter

- Sprint set to Sprint 4

#19 - 03/08/2018 07:57 PM - bmbouter

- Sprint/Milestone deleted (22)

#20 - 04/15/2019 10:28 PM - bmbouter

- Tags Pulp 2 added