RPM Support - Issue #1786

comps.xml has doctype with invalid dtd reference

03/21/2016 03:11 AM - rmcgover

Status:	CLOSED - WONTFIX		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:			**Estimated time:**	0:00 hour
Category:				
Sprint/Milestone:				
Severity:	1. Low		**Groomed:**	No
Version:	Master		**Sprint Candidate:**	No
Platform Release:			**Tags:**	Pulp 2
OS:			**Sprint:**	
Triaged:	Yes		**Quarter:**	

Description

yum distributor's generated comps.xml files contain the doctype declaration:

```
<!DOCTYPE comps PUBLIC "-//Red Hat, Inc.//DTD Comps info//EN" "comps.dtd">
```

Attempting to fetch that comps.dtd URI will fail (there's no such file provided relative to the comps.xml file or anywhere else as far as I can tell).

This probably doesn't matter too much but I've filed this bug mainly for this reason: the python 3 XML SAX parser in its default configuration will attempt to fetch external entities, and will crash since comps.dtd can't be loaded.

Other tools, such as xmllint --loaddtd, will also complain.

History

#1 - 03/24/2016 08:43 PM - mhrivnak

- Triaged changed from No to Yes

#2 - 03/24/2016 08:44 PM - mhrivnak

- Parent task set to #1683

#3 - 04/25/2016 03:15 PM - bmbouter

The comps.xml does specify an invalid DTD reference, but I'm not sure what to replace it with.

#4 - 04/29/2016 09:01 PM - bmbouter

After e-mailing a packaging mailing list I've gotten several possible ones. I've looked over them and I think this one is the most complete:
https://git.fedorahosted.org/cgit/comps.git/tree/comps.dtd

This one defines all 4 types I have seen in comps.xml files including (group, category, langpacks, and environment).

    https://git.fedorahosted.org/cgit/comps.git/tree/comps.dtd

This one is missing definitions for environment, and langpacks, but is otherwise identical to ^

    A DTD file is contained in the comps-extras package which installs it here: /usr/share/doc/comps-extras/comps.dtd

Also this project. It is a C library and I don't think we want to pick up a new dependency to solve this issue.

    libcomps [1] project was created for the purpose to set a standard but I am afraid that there's no DTD. So if XML is parsable with libcomps then it should be valid.

    [1] https://github.com/midnightercz/libcomps

This one only defines category and groups so it is also not as expressive as the top most one. Also it is not a DTD file.

there was a relax NG specification made: https://github.com/rpm-software-management/yum/blob/master/docs/comps.rng

#5 - 06/13/2016 04:39 PM - jluza

I would stick with official fedora DTD https://pagure.io/fedora-comps/blob/master/f/comps.dtd

#6 - 06/13/2016 07:50 PM - bmbouter

jluza, Thanks your recommendation seems great

#7 - 06/29/2016 04:21 PM - bmbouter

- Sprint Candidate changed from No to Yes

#8 - 06/29/2016 06:57 PM - jcline@redhat.com

I'm in agreement it should be fixed, but I think it needs to be fixed in the upstream tooling (probably just createrepo_c). It appears the latest Fedora composes have the same doctype declaration[0].

[0] http://ftp.linux.ncsu.edu/pub/fedora/linux/releases/24/Server/x86_64/os/repodata/

#9 - 06/29/2016 07:33 PM - bmbouter

I agree jcline. I've filed the upstream bug here: https://github.com/rpm-software-management/createrepo_c/issues/62

#10 - 05/08/2017 08:12 PM - bmbouter

- Tags RCM added

#11 - 09/14/2018 03:57 PM - amacdona@redhat.com

- Sprint Candidate changed from Yes to No

#12 - 04/12/2019 10:17 PM - bmbouter

- Status changed from NEW to CLOSED - WONTFIX

Pulp 2 is approaching maintenance mode, and this Pulp 2 ticket is not being actively worked on. As such, it is being closed as WONTFIX. Pulp 2 is still accepting contributions though, so if you want to contribute a fix for this ticket, please reopen or comment on it. If you don't have permissions to reopen this ticket, or you want to discuss an issue, please reach out via the developer mailing list.

#13 - 04/15/2019 10:33 PM - bmbouter

- Tags Pulp 2 added

#14 - 05/03/2019 04:05 PM - bmbouter

- Tags deleted (RCM)