Infrastructure - Bug #3492

Invalid PIDs in production (whitespace)

2013-01-17 15:13 - Dave Vieglais

Status:	In Progress		**Start date:**	2013-01-18
Priority:	Urgent		**Due date:**	
Assignee:	Jing Tao		**% Done:**	30%
Category:	Support Operations		**Estimated time:**	0.00 hour
Target version:	Maintenance Backlog			
Milestone:	None		**Story Points:**	
Product Version:	*			

Description

Recording this for future reference.

There are nine PIDs in the production environment that contain whitespace. This appears to have no functional effect - sysmeta and objects can be retrieved so no action is required other than to ensure no more sneak in.

The PIDs in question are:

guid

doi:10.5063/AA/Virginia Perez.2.1
judith kruger.3.1
judith kruger.4.1
judith botha.1.1
judith kruger.1.1
judith kruger.2.1
judith kruger.5.1
judith botha.2.1
resourceMap_Lin Cheng-Jung.1.1
resourceMap_Lin Cheng-Jung.1.2
resourceMap_Lin Cheng-Jung.1.3
Lin Cheng-Jung.1.1
Lin Cheng-Jung.1.2
Lin Cheng-Jung.1.3
doi:10.6085/AA/ SHLX00_XXXITV2XLSR03_20111128.40.1

Subtasks:

Task # 3498: Obsolete these invalid PIDs on SANParks	**Closed**
Task # 3499: Investigate Virgina Perez.2.1 object	**Closed**
Task # 3131: Whitespace in pisco PID	**Closed**
Bug # 4674: Ask Judith, Mike and Virgina Perez.2.1 to obsolete those pids which contain...	**New**
Task # 4675: Administratively delete those pids with the white spaces	**New**

History

#1 - 2013-01-17 15:17 - Dave Vieglais

There are nine PIDs in the production environment that contain whitespace.

Actually eight, one dupe in the cut'n paste:

doi:10.5063/AA/Virginia Perez.2.1
judith botha.1.1
judith botha.2.1
judith kruger.1.1
judith kruger.2.1

judith kruger.3.1
judith kruger.4.1
judith kruger.5.1

#2 - 2013-01-18 17:41 - Ben Leinfelder

- Description updated

#3 - 2013-06-20 20:41 - Dave Vieglais

Some new invalid PIDs appeared in production, mostly from the TFRI node:

Lin Cheng-Jung.1.2
Lin Cheng-Jung.1.3
Lin Cheng-Jung.1.1
resourceMap_Lin Cheng-Jung.1.1
resourceMap_Lin Cheng-Jung.1.3
resourceMap_Lin Cheng-Jung.1.2
doi:10.6085/AA/ SHLX00_XXXITV2XLSR03_20111128.40.1

#4 - 2013-08-14 18:11 - Dave Vieglais

- Assignee changed from Dave Vieglais to Robert Waltz

#5 - 2013-10-28 20:01 - Robert Waltz

- Product Version set to 1.3.0

- Milestone changed from None to CCI-1.3

#6 - 2013-10-28 20:45 - Robert Waltz

- Milestone changed from CCI-1.3 to CCI-1.2

- Target version set to 2013.44-Block.6.1

- Due date set to 2013-11-09

- Product Version changed from 1.3.0 to *

- Start date set to 2013-10-27

#7 - 2014-01-06 18:35 - Chris Jones

- Assignee changed from Robert Waltz to Ben Leinfelder

#8 - 2014-01-06 18:36 - Chris Jones

- Due date changed from 2013-11-09 to 2014-02-01

- Target version changed from 2013.44-Block.6.1 to 2014.4-Block.1.2

#9 - 2014-01-06 18:36 - Chris Jones

- Assignee changed from Ben Leinfelder to Jing Tao

#10 - 2014-02-12 17:08 - Jing Tao

- Status changed from New to Closed

We added the code in metacat to enforce there is no white spaces in the docid.

#11 - 2014-02-12 17:52 - Jing Tao

- Status changed from Closed to In Progress

Reopen the bug - we still need to contact the administrator to ask them to change the docids.

#12 - 2014-03-14 17:49 - Chris Jones

- Target version changed from 2014.4-Block.1.2 to 2014.12-Block.2.2

- Due date changed from 2014-02-01 to 2014-03-29

#13 - 2014-04-01 22:59 - Chris Jones

- Description updated

#14 - 2014-09-02 21:42 - Robert Waltz

- Due date changed from 2014-03-29 to 2014-09-02

- Milestone changed from CCI-1.2 to None

- Target version changed from 2014.12-Block.2.2 to Release Backlog

#15 - 2014-09-24 18:09 - Dave Vieglais

- Due date changed from 2014-09-02 to 2014-09-24

- Target version changed from Release Backlog to Maintenance Backlog

#16 - 2016-11-30 18:27 - Rob Nahf

- % Done changed from 20 to 30

the recent upgrade to 2.3 changes the behavior somewhat for retrieval, due most likely to JAXB schame validation which replaced our custom JiBX validation.

The main difference is that a /meta call returns a ServieFailure. A /cn/object call will still return the object (if it's on the CN) if you unicode escape the space on the URL with %20.

Also, listObjects will omit the object from the list but return other pids.

"Lin Cheng-Jung.1.3" is a pid with no access policy, but via psql, I found the modified date for it, and did a time-bound slice of listObjects:

https://cn.dataone.org/cn/v2/object?fromDate=2013-06-17T03:31:00.000Z&toDate=2013-06-17T03:32:00.000Z