

## Infrastructure - Task #1142

Bug # 1141 (Closed): unicode guid handling in mn.create()

### address MMP errors for supplementary unicode characters in pids

2010-12-03 21:28 - Rob Nahf

<b>Status:</b>	Rejected	<b>Start date:</b>	2010-12-03
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Robert Waltz	<b>% Done:</b>	0%
<b>Category:</b>	d1_cn_service	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>Story Points:</b>	
<b>Milestone:</b>	None		
<b>Product Version:</b>			
<b>Description</b>			
the error message returned from the test for guid string: common-unicode-supplementary-escaped-			
Error creating MMP stream in MNode.handleCreateOrUpdate: Could not serialize the system metadata to multipart: Illegal character code 0xd800 in content text.			
The error happens within the mctx.marshalDocument(object, "UTF-8", null, out) call in serializeServiceType() in D1Node.java.			
This error doesn't happen with other unicode characters in the BMP.			

### History

#### #1 - 2010-12-15 20:49 - Rob Nahf

Looks like a byteStream error of not using UTF-8

unicode String:: common-unicode-supplementary-escaped-

#### #2 - 2010-12-15 23:02 - Rob Nahf

- Assignee set to Robert Waltz

#### #3 - 2010-12-17 23:46 - Rob Nahf

- Priority changed from Urgent to Normal

not as urgent since metacat is not going to support non-ascii characters for a while. (this bug only shows itself with supplementary characters beyond the BMP range)

#### #4 - 2011-02-04 17:12 - Robert Waltz

- Milestone set to CCI-0.5

- Category set to d1\_cn\_service

#### #5 - 2011-03-25 17:00 - Robert Waltz

- Milestone deleted (CCI-0.5)

#### #6 - 2011-05-16 17:57 - Dave Vieglaiss

- Target version changed from Sprint-2011.18-Block.3 to Sprint-2011.20-Block.3

#### #7 - 2011-05-23 17:38 - Robert Waltz

- Target version deleted (Sprint-2011.20-Block.3)

this is an issue with the way jibx handles these characters. If we wish to fix this for the long run, we'll probably have to submit a solution to the jibx project.

**#8 - 2012-02-10 03:47 - Robert Waltz**

- Milestone set to CCI-1.0.0

- Status changed from New to In Progress

The JibX IXMLReaderFactory implementation, XMLPullParserFactory, uses an InputStreamWrapper class that contains a UTF8Reader ( WrappedStreamUTF8Reader) unable to support Surrogate pairs.

The JibX runtime can override the default IXMLReaderFactory if calling code of the Unmarshaller sets a org.jibx.runtime.impl.parser System Property.

I have identified a UTF8StreamReader class that supports Surrogate pairing (and hence the UTF supplementary character range) in a separate library: javolution.io.UTF8StreamReader. It is open source with a very liberal copy right. I can create modify the JibX implementation of XMLPullParserFactory so that is uses a modified InputStreamWrapper incorporating a copy of the javolution.io.UTF8StreamReader.

The above may solve the problem if the XmlPull dependencies do not blow up on surrogate pairs.

**#9 - 2012-06-19 20:05 - Robert Waltz**

- Milestone changed from CCI-1.0.0 to CCI-1.1

**#10 - 2012-09-26 17:05 - Robert Waltz**

- Milestone changed from CCI-1.1 to CCI-1.2

**#11 - 2013-03-01 18:37 - Robert Waltz**

- Subject changed from address MMP errors for supplementary unicode characters in guids to address MMP errors for supplementary unicode characters in pids

**#12 - 2014-06-30 22:30 - Robert Waltz**

- Milestone changed from CCI-1.2 to None

**#13 - 2016-08-10 20:37 - Robert Waltz**

- translation missing: en.field\_remaining\_hours set to 0.0

- Status changed from In Progress to Rejected

We no longer use JibX