

## EDIT - feature request #7800

### Parse preliminary RefDetails

09/29/2018 12:30 PM - Andreas Müller

<b>Status:</b>	Closed	<b>Start date:</b>	
<b>Priority:</b>	Highest	<b>Due date:</b>	
<b>Assignee:</b>	Andreas Müller	<b>% Done:</b>	100%
<b>Category:</b>	cdmadapter	<b>Estimated time:</b>	5:00 hours
<b>Target version:</b>	Euro+Med Migration		
<b>Severity:</b>	normal		
<b>Description</b>			
There are >40000 RefDetails imported as the are preliminary. We could try to parse them during import.			
<b>Related issues:</b>			
Related to EDIT - feature request #7799: AM: Parse authorteams			<b>Resolved</b>
Related to EDIT - feature request #7801: AM: Deduplicate references			<b>Closed</b>
Related to EDIT - bug #7829: Improve deduplication of parsed names and refere...			<b>Closed</b>

#### Associated revisions

##### Revision 14ec817b - 10/10/2018 12:45 PM - Andreas Müller

ref #7800 parse preliminary RefDetails (first start)

##### Revision 517ecf1f - 10/16/2018 05:52 PM - Andreas Müller

ref #7800 add specific subclass matching to matching strategies

##### Revision 1c4a145c - 10/16/2018 06:05 PM - Andreas Müller

ref #7829 ref #7800 remove nomenclaturallyRelevant from Reference matching as it is not used at all

##### Revision 8643e56e - 11/08/2018 10:19 AM - Andreas Müller

ref #7429 fix order in "publ." parsing and implement for nom. ref. parser

##### Revision 62e62a2f - 11/08/2018 10:26 AM - Andreas Müller

some improvements to NonViralNameParser

##### Revision a258603e - 11/08/2018 10:47 AM - Andreas Müller

fix parser test

##### Revision f6687505 - 11/08/2018 11:56 AM - Andreas Müller

add regEx negation to UTF8

##### Revision 473b1c68 - 11/08/2018 11:57 AM - Andreas Müller

Add series parsing with letters

##### Revision 943941be - 11/08/2018 12:01 PM - Andreas Müller

add parsing of brackets with "&"

##### Revision fbe910aa - 11/08/2018 12:20 PM - Andreas Müller

ref #7829, ref #7800 improve nom. ref. matcher (also changes the result type => needs adaptation in client Apps)

##### Revision 88710aed - 11/09/2018 03:02 PM - Andreas Müller

ref #7800 improve parsing of details and reference titles

##### Revision e6883603 - 11/09/2018 03:50 PM - Andreas Müller

ref #7800 fix tests with special characters

**Revision a0976016 - 11/09/2018 04:33 PM - Andreas Müller**

ref #7800 2nd fix tests with special characters

**Revision 0c5910cb - 11/09/2018 08:34 PM - Andreas Müller**

ref #7800 further parsing for volume and edition part and for some special title parts

## History

---

**#1 - 09/29/2018 12:31 PM - Andreas Müller**

- Related to feature request #7799: AM: Parse authorteams added

**#2 - 09/29/2018 12:32 PM - Andreas Müller**

- Related to feature request #7801: AM: Deduplicate references added

**#3 - 10/15/2018 03:45 PM - Andreas Müller**

- Status changed from New to In Progress

- Priority changed from New to Highest

**#4 - 10/16/2018 06:02 PM - Andreas Müller**

- Related to bug #7829: Improve deduplication of parsed names and references added

**#5 - 12/18/2020 11:37 PM - Andreas Müller**

- Status changed from In Progress to Closed

As the import did run and most of the above RefDetails were parsed I think we can close this ticket. However, there are still >1000 unparsed references (I did already spent some time for cleaning them up where possible). Further work needs to be done by ERS or references will be removed when removing Kew data.

**#6 - 12/18/2020 11:37 PM - Andreas Müller**

- % Done changed from 0 to 100