

Infrastructure - Decision #577

Define maximum length of identifiers

2010-04-13 19:19 - Dave Vieglais

Status:	Closed	Start date:	
Priority:	High	Due date:	
Assignee:	Dave Vieglais	% Done:	100%
Category:	Documentation	Estimated time:	0.00 hour
Target version:			
Milestone:	None		

Description

The identifier type is currently defined as "unicode string of printable characters, excluding whitespace". There are however, practical limitations to the maximum length of strings imposed by the various uses and storage of identifiers.

URLs have a practical length limit of about 2048 characters (primarily imposed by IE). Apache ups this to about 4000.

In data stores, identifiers will likely be used as primary keys. The limitations imposed by various relational DBs include:

[MySQL](#): CHAR or VARCHAR = 256

Postgres: 1e9 chars (2GB storage limit)

sqlite: 1e9 (configurable)

URLs are often used as identifiers, and these can be quite long (certainly longer than 256 chars), so 256 seems like an unreasonable limitation, and technical solutions are available for this (e.g. in mysql store identifier as text and create a lookup table with a hash of the identifier as key).

Identifiers are used in [DataONE](#) REST calls (e.g. get() and resolve()), so the total length of the [DataONE](#) URL should not exceed the practical URL length limitations imposed by common browsers and libraries.

So, given that IE apparently imposes a 2k limit on URLs, split that in half to 1k.

Evaluation of identifiers in KNB by Matt indicates a maximum identifier length of 239 characters.

Dryad IDs are a max of 38, extending soon to 60 chars.

At the Feb Santa Barbara we came up with a max length of 800 chars, which seems kind of arbitrary, though safe from a functional perspective.

So 800 chars it is. Updating the Types docs.

History

#1 - 2010-04-15 22:20 - Dave Vieglais

- Status changed from New to Closed

Docs were updated.