

TET Migration Guide

Latest version covered in this document: TET 5.3

The TET API generally remains compatible among major releases. Sometimes API methods are phased out and replaced with an improved successor. In these situations a newer version declares the old method as deprecated without changing the functionality in any way. In most cases the new method is more general or more powerful than the old one which it replaces. The new method is always preferable. New code should not use deprecated API methods, and existing code should switch to the new replacement method. Deprecated methods may be removed in a future release.

This document contains recommendations for users who migrate existing TET application code which has been developed with an older release. The migration guide explains how to identify deprecated API features which are used in application code. Once identified, the deprecated features should be replaced with the recommended newer ones.

Note The document compatibility.txt in the distribution package contains additional version-by-version information regarding API compatibility.

1 Identify deprecated TET API Features

1.1 Identify deprecated TET Method Calls at Compile Time

Since TET 5.2 and TET 5.3 don't contain any API methods which are marked as deprecated this section does not apply.

1.2 Identify deprecated API Method Calls at Runtime

Since TET 5.2 and TET 5.3 don't contain any API methods which are marked as deprecated this section does not apply.

1.3 Identify deprecated Options

All language bindings. Deprecated options can be identified in the logging output with the logging class `api=2` which can be set as follows:

```
tet.set_option("logging={filename=deprecated.log classes={api=2}}");
```

With this setting the log file contains a message for all API methods which are called with deprecated options, e.g.

```
TET_open_page(tet_0x7ffffba9c0e70, 0, 3, "imageanalysis={smallimages={maxheight=50}}")  
[Option "smallimages" is deprecated]
```

These logging entries are only warnings. TET execution continues and PDF contents are extracted as usual. However, it is strongly recommended to adjust the application code to avoid deprecated options.

2 Deprecated TET API Features

2.1 Deprecated TET Methods

Table 2.1 *Deprecated TET API functions/methods and recommended replacements*

deprecated API method	deprecated since	removed in	recommended replacement method
<code>TET_document_mem()</code>	TET 3.0	TET 5.1	<code>TET_create_pvf()</code> and <code>TET_open_document()</code>
<code>TET_get_xml_data()</code>	TET 5.0	TET 5.2	<code>TET_get_tetml()</code>
<code>TET_utf16_to_utf8()</code>	TET 4.1	TET 5.2	<code>TET_convert_to_unicode()</code>
<code>TET_utf8_to_utf16()</code>			
<code>TET_utf32_to_utf8()</code>			
<code>TET_utf8_to_utf32()</code>			
<code>TET_utf16_to_utf32()</code>			
<code>TET_utf32_to_utf16()</code>			

2.2 Deprecated Options

Table 2.2 *Deprecated TET API options and recommended replacements*

deprecated option or suboption	deprecated since	removed in	recommended replacement option
<code>TET_set_option()</code>:			
filenamehandling, keyword legacy	TET 5.3	TET 5.3	use explicit encoding name or honorlang
<code>TET_open_document()</code>:			
tetml, suboption version	TET 4.0	TET 4.4	n/a (TET 3 XML output is no longer supported)
tetml, suboption elements, suboption docxmp	TET 5.0	TET 5.3	suboption metadata
keepua	TET 4.0	TET 5.3	<code>TET_open_document()</code> , option fold with one the following foldings: fold={{[:Private_Use:] preserve}} or fold={{[:Private_Use:] unknownchar}}
<code>TET_open_page()</code> and <code>TET_process_page()</code>:			
imageanalysis, suboption smallimages	TET 5.2		suboptions heightrange, sizerange, widthrange
skipengines	TET 5.0	TET 5.3	engines
contentanalysis, suboption ideographic	TET 5.0	TET 5.3	n/a (the default is now the reasonable behavior of ideographic=keep)
contentanalysis, suboptions lineseparator, paraseparator, wordseparator	TET 4.0	TET 5.3	options lineseparator, paraseparator, wordseparator of <code>TET_open_document()</code>
contentanalysis, suboption zoneseparator	TET 3.0	TET 5.2	n/a (option was never supported and didn't have any effect since TET 5.0)
<code>TET_write_image_file()</code> and <code>TET_get_image_data()</code>:			
compression, preferredtiffcompression	TET 5.2		n/a

Table 2.2 Deprecated TET API options and recommended replacements

<i>deprecat</i> ed option or suboption	<i>deprecat</i> ed since	<i>removed in</i>	<i>recommended replacement option</i>
format	TET 4.0	TET 5.2	n/a (didn't have any effect since TET 4.0)
smallimages, suboption maxcount	TET 5.0	TET 5.3	n/a

3 Incompatible API Changes

Table 3.1 Incompatible TET API changes and required changes

API change	changed in	required change
<i>TET_new2()</i> : the errorhandler callback no longer uses the errortype parameter which has been unused since TET 4	TET 5.3	Remove the errortype parameter from your custom errorhandler

4 TETML Changes

Table 4.1 lists changes in the TETML output which may affect post-processing with XSLT or other means. Changes which merely add elements, attributes or attribute values in a compatible way are not listed (see version information for elements and attributes in the TET manual).

Table 4.1 Changes in TETML which may affect post-processing

TET version	relevant change	notes
TET 5.3	<Glyph> element	Previously ligatures with a non-private Unicode value were already resolved when the <Glyph> element was written or output with granularity=glyph was created, i.e. the corresponding sequence was emitted regardless of the decompose option. Now ligatures with a Unicode value are preserved in the <Glyph> element. In higher-level TETML elements they are processed according to the decompose option.
TET 5.1	new elements	The new elements List, Item, Label, and Body have been introduced to describe the results of list detection, but they are not created by default. List detection can be enabled with the page option structureanalysis={list=true}.
TET 5.0	namespace URI	The URI for the xmlns:tet namespace has been changed to reflect the new schema.
TET 5.0	additional Box elements	The Para and Table elements now have additional Box elements as child.

5 TET History

Table 5.1 TET release history

TET version	release date
TET 3	2009
TET 4	2010
TET 4.1	2012
TET 4.2	2013
TET 4.3	2014
TET 4.4	2015
TET 5	2015
TET 5.1	2017
TET 5.2	2019
TET 5.3	2021