



ADELARD

# Linking to PDF

## Plugin Manual for ASCE 4.2

---

- File name: linking to pdf v0.1.xml
- Version: 0.1.24
- Applies to: All networks

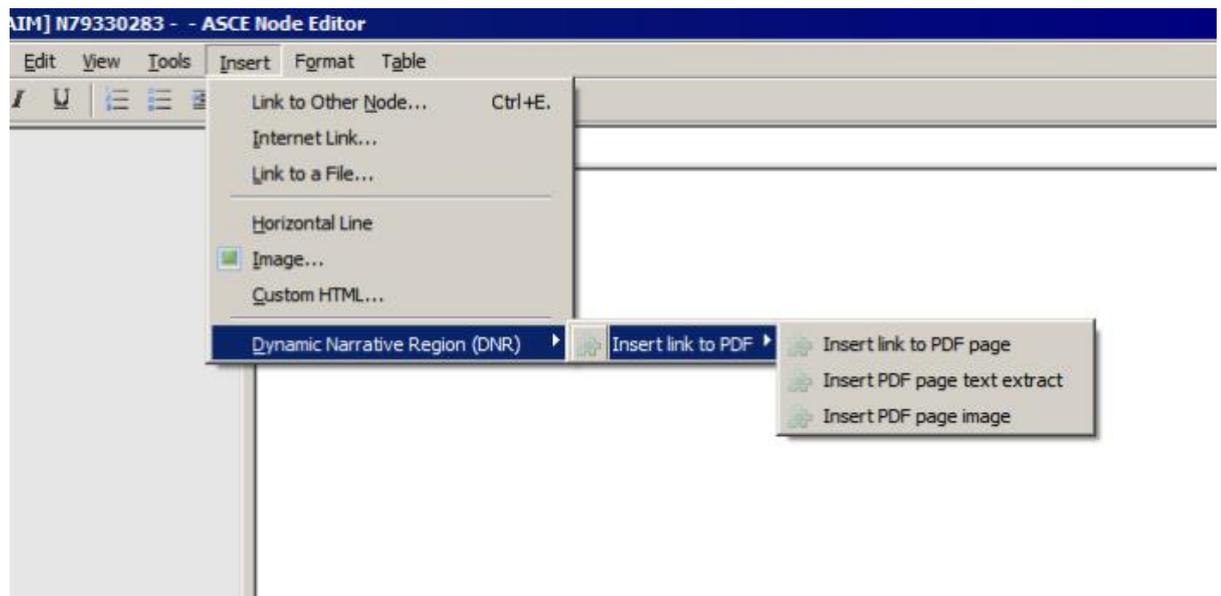
### 1 Description

This plugin provides three DNRs for linking to or extracting content from PDF files, using the open source [XPDF](#) PDF utilities.

It also provides a deep navigation feature to specific pages of PDF documents, in recent versions of Adobe Acrobat.

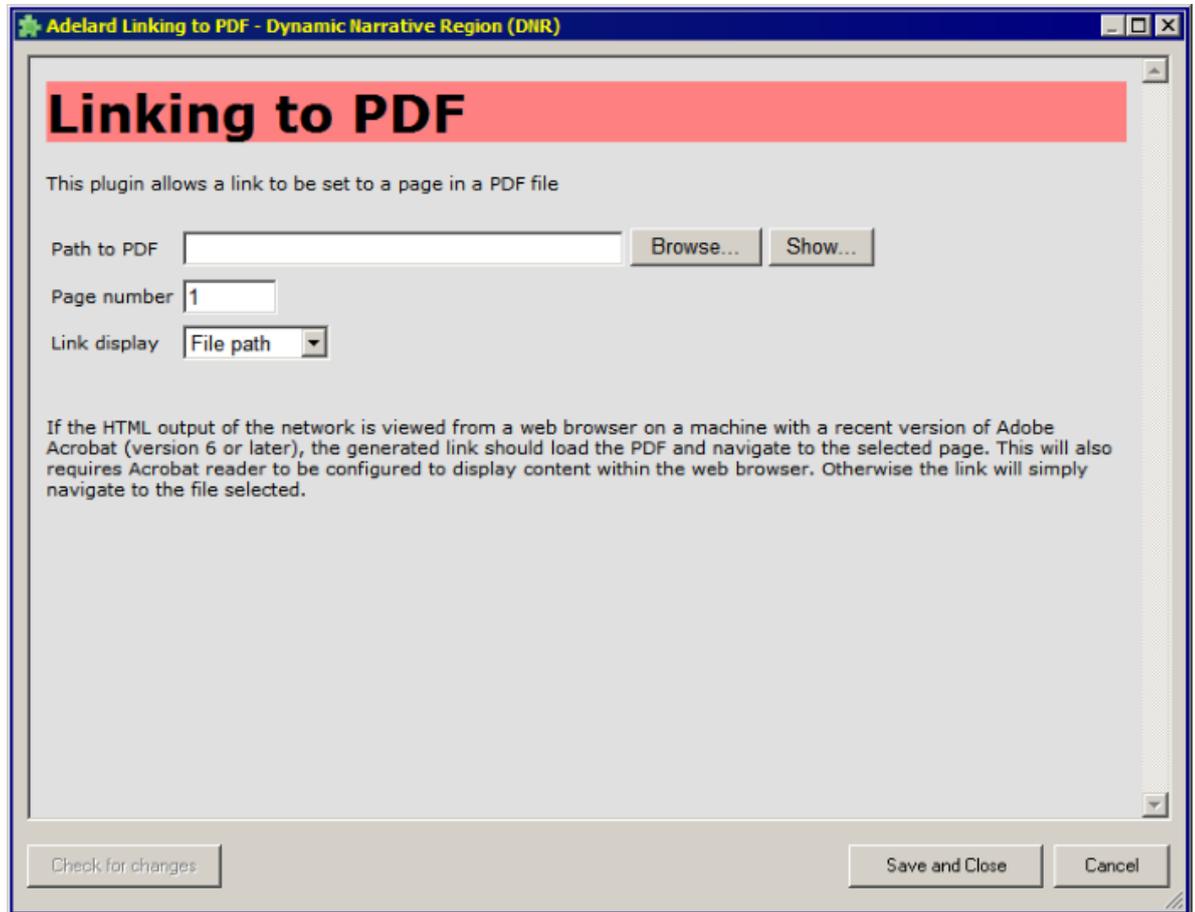
This plugin works with the one click export plugin "remove DNR containers" feature.

For times when you need to reference material from PDFs, this plugin provides you with three choices. These are described further below.



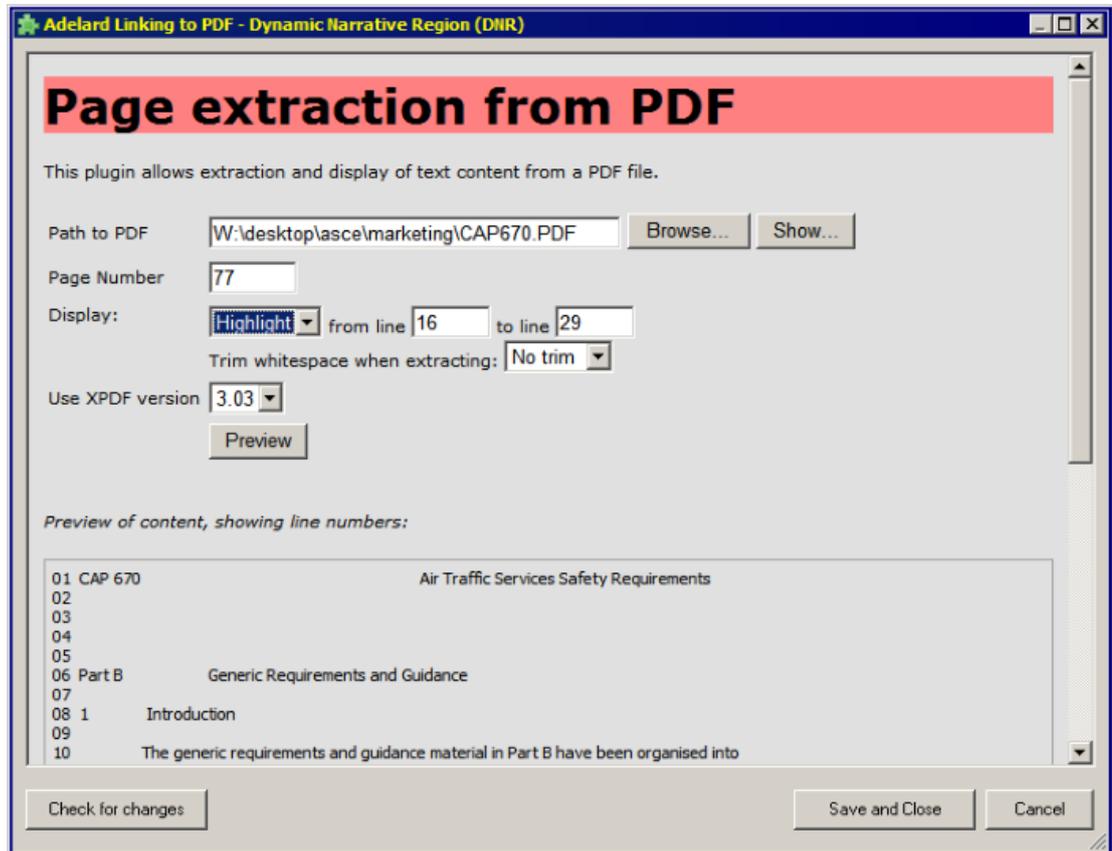
### 2 Link to a page

Create a link to a PDF file, specifying the page number you'd like the file to open at when you click on the link



### 3 Extract from a page

Extract text from a PDF file and embed it into your node text as a Dynamic Narrative Region (DNR). Note that extracts are formatted a plain text only. As with all ASCE DNRs, changes to the remote "source" file can be easily detected and viewed.



### 3.1 Extraction options

- highlight - marks the specified lines with a yellow background like a highlighter pen. This option is useful when you want to draw attention to certain lines on the page, but enable them to be seen in the context of the rest of the page.
- extract - extracts the specified lines only.

White space can be left-trimmed from the extracted or highlighted text.

XPDF version - specifies which version of XPDF to use to extract the text from the chosen file. See the compatibility aspects below to understand when you might use an older version of XPDF.

---

### 3.2 Extract from a page - compatibility aspects relating to XPDF

- An earlier version of this plugin used XPDF version 3.02. This older version may experience issues in extracting from recent PDF files, particularly those containing unicode, or that are PDF 1.6 files. As a result the latest version of this plugin uses XPDF version 3.03 as the default.
- XPDF 3.03 has a slightly different layout algorithm on some PDF files, so we have decided not to automatically upgrade existing DNRs to use XPDF 3.03 automatically, as existing DNRs may therefore become invalid in virtue of the content being offset. There is now an option to choose which version of XPDF to use in extracting page texts. We recommend using XPDF 3.03.
- Existing DNRs created with an earlier version of this plugin are set to use the XPDF 3.02 version
- As a result, there is a possible compatibility issue if DNRs created with the latest version of this plugin are configured to use PDF 3.03, and then the content is shared with a user who only has an older version of the plugin. When they refresh the DNRs, they will be using XPDF 3.02 and in a few cases the extractions will not be exactly the same as that seen in the current version.
- We recommend if you are planning to exchange content with users that have a previous version of this plugin that you either:
  1. Ask them to install this later version of the plugin, or
  2. Continue to use XPDF 3.02 for your PDF extracts

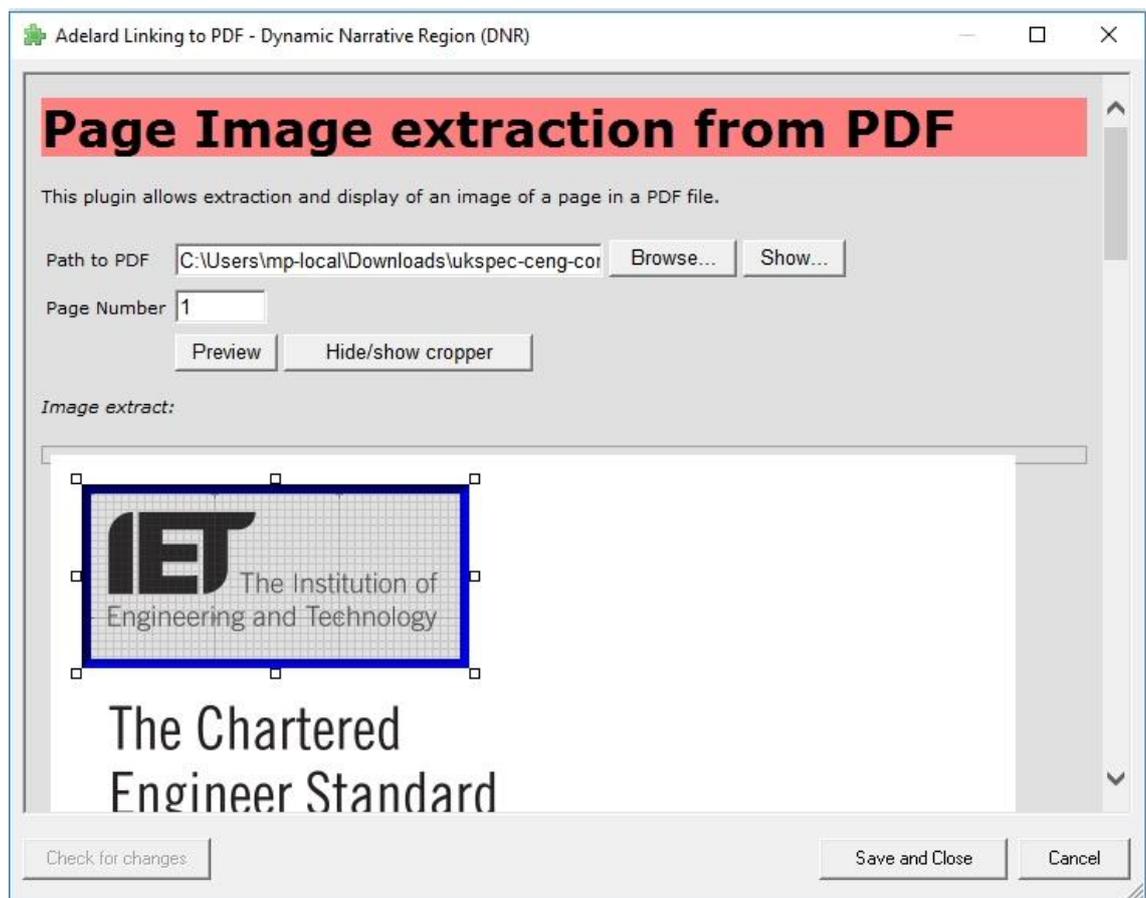
### 4 Insert PDF Page Image

This option allows you to insert the image of a page from a PDF file. It is similar to taking a screenshot of a page of a PDF file and then embedding it in your ASCE network. This DNR provides a simple interface to select a PDF file and the page to be extracted. Because the content is extracted as an image, it will work better with PDFs that are not so amenable to text based extraction due to:

- Significant use of complex text layout or tables
- Images and diagrams

- OCR (Optical Character Recognition) based PDFs, such as those created by scanning paper documents
- PDFs where a more visual extraction is desired.

The following screenshot shows the DNR editor window previewing an image of a page in a PDF file.



It's also possible to insert only a sub-section of an image by using a cropping box provided by the plugin. To enable the cropper simply click on the "Hide/show Cropper" button and position the cropping box so that it encompasses the area you want to keep. In the picture above, the cropping box is the blue box displayed on top of the picture preview. Once you have set the cropper click "save and close", the picture will be cropped to the selection you made and only the area within the boundaries of the box will be inserted, as a new picture, in the DNR content.

Please note that the crop box cannot be moved directly, to move it in a different area of the picture you'll have to resize the appropriate sides of the box. This new image will also be visible in the ASCE global list of images (View -> Images) and will be preserved if the network is exported to a MS Word file.

## 5 FAQ

*Q: Why when I press the "show" button do I sometimes get a different page in Acrobat to the one specified for certain files?*

A: Some PDF files have preliminary pages before Page 1 (often numbered with Roman numerals such as *i*, *ii*, *iv* and so on). For example if you want to navigate to Page 3, when the text is extracted, the third page in the file will be extracted - this is the behaviour of XPDF. However when the plugin fires up Acrobat to navigate to the specified file, Acrobat will navigate to the page called "Page 3", which may not actually be the third page in the file, but rather the third page *after* the preliminary pages. We do not have a solution to this issue at present.