Archiving a single .pdf file posted online using the 'save page' functionality offered by the Internet Archive

Until June 2015, access through cataloging records to free electronic resources was through a link to the live site only. It has not been possible for staffing reasons to maintain access effectively over time as content changes or links break. In 2015, the Internet Archive began to offer a free service that can reliably capture monographic PDFs.

The procedure below gives instructions for selectors to create an archived version of a PDF in the Internet Archive Wayback Machine and submit that link (as well as the live site link) to catalogers by means of the Internet Resource Cataloging Request Form (IRCR) The archived version and the live site version will be identical which could be useful for dating purposes or to assure citation integrity. This extra step will also preserve access to content even if the live site disappears.

Note: Only **single volume monographs in PDF** can be archived using this procedure.

Procedure:

In order to capture a single pdf posted on a website, use the "Save Page Now" functionality via the Internet Archive:

1. Navigate to the pdf you would like to save and then copy the URL. To illustrate the following steps, we used this URL: http://www.locjkt.or.id/application/docliblcjkt/upload/ldbMP2020Frmwrk.pdf

2. Go to https://archive.org/web and look to the lower right corner of the screen for "Save Page Now". Paste the URL you are submitting into the box and click 'Save page'.

3. The pdf will be saved and you will be given a permanent URL for the pdf you just archived. Check to make sure the pdf was captured fully and copy the link. (i.e. http://web.archive.org/web/20150512202520/http://www.locjkt.or.id/application/docliblcjkt/upload/ldbMP2020Frmwrk.pdf)

4. Copy and paste the link to the archived pdf in the Archived Site box on the IRCR form.

There are a few caveats and points for fair warning about saving pdfs using these methods:

1. You will be capturing a single pdf, not an entire site or the series of pdfs linked to from a web page. Other pages might be available in the Internet Archive and be accessible when a user views the page you archived but what's captured can be very hit or miss on large web crawls
2. Some pdfs might not be capturable because the website they are on may block web crawlers and prevent them from being captured (for more on blocking crawlers please see: http://www.robotstxt.org/robotstxt.html)
3. The archived copy might not be a perfect copy of the original (but with pdfs it should be possible to capture the full pdf)
4. If you cannot capture the pdf you can search for it by entering the URL into the Wayback Machine's search box (the search box is at the top of this page: https://archive.org/).

Step by step guide to capturing a pdf file posted online:

Go to https://archive.org/web and look to the lower right corner of the screen for "Save Page Now". Paste the URL you are submitting (i.e. http://www.locjkt.or.id/application/docliblcjkt/upload/ldbMP2020Frmwrk.pdf) into the box, check the full URL has been inserted and click 'Save page'.
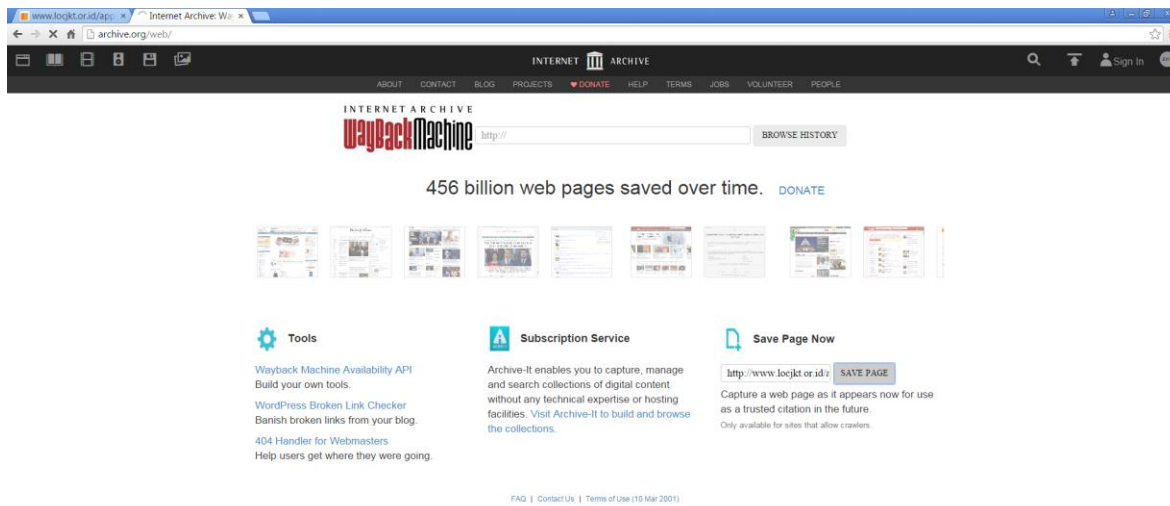
The link generated will be the link to the archived version of the pdf. In the URL for the archived pdf there will be some information including the date and time the pdf was saved and the full original URL (i.e. http://web.archive.org/web/20150512202520/http://www.locjkt.or.id/application/docliblcjkt/upload/ldbMP2020Frmwrk.pdf). When archiving a pdf file you might get a dialog box containing the URL for the archived page but often you will just see the archived pdf and its associated URL in a new tab or window:
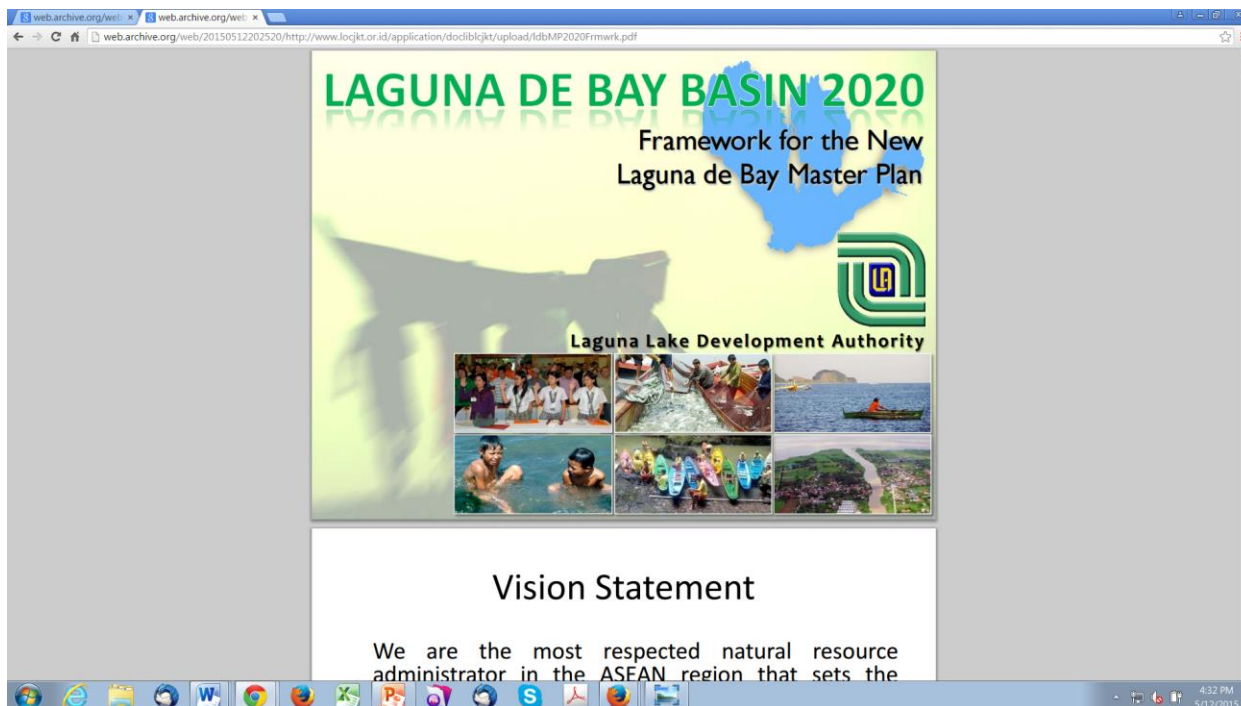
Scroll down to make sure the full pdf was captured:

alp/kmh 5/28/2015