Procedure for Creating an Archive for Magazines in PDF Format in the Internet Archive (Archive.org) by Gardner Patton

First format all the issues of your magazine in .PDF format with one issue per file. Rename the files properly for upload. Each file should have the name of the magazine (separate words with underlines) and the year and month of publication. If you don't have a year and month use the Vol and No.. If you don't have that, number them sequentially. Thus the file name would be "magazine_name_yearmonth.pdf or "magazine name volxxserxx.pdf . The yearmonth or volser sequence allows for a date order sort of the files.

Now that the files are organized and named there are 2 ways to upload them. One way is to load each one as a single item with one magazine in each item. The second way is to have a single item (which will become a collection of issues) with each .PDF as a file within that item. Note we are assuming a .PDF file, not individual .jpg images which is often the case when an item represents one issue.

The first method is the more traditional way and there is meta data for each issue which contains among other things the issue date of the magazine and the number of pages. Also with this method the first page of the issue is preserved and shown by the reader. With this method however the meta data must be supplied for each item. This can be speeded up by using parameters in the command used to invoke the uploader page but that information may need to be modified slightly for each issue (date changes, etc.).

Using the second method the meta data is only entered once for all the issues. Thus you lose the opportunity to enter the creation date of the issue and the number of pages. However, if you have named the files with the creation date, that date is easily available to viewers. The number of pages is rarely significant. With the second method you can queue up hundreds of files at a time for upload. The upload times for each file seem about as fast as FTP. Using the procedure to add files to an existing item (described below), an upload of a large number of issues can be broken into manageable sessions. However, if you upload hundreds of large files at one time the Archive may take many days to process them all (create derivative files). In the meantime they are available as normal .PDF files with the permalinks assigned as they are uploaded.

For these reasons I prefer to use the second method for uploading multiple .PDF files each containing one issue of a magazine, a whole collection of the same magazine in one or several uploads, but not one upload for each issue.

1) To do this log in the Archive and click the Upload button at the top of the page. See image on next page.



2) Click Upload Files.



3. Drag 1 or more files to the pink box that you want to upload.

) 🔒 https://arc	hive.org/upload/	es to the internet archive 🔶 👃 🍙 🏠		e
t Visited 👩 Wri	iting and Illustratin 😚 https://www.facebook 🦎 Diana Patton Ar	sst 🧬 New Square Music - h. 🛞 Bee Sh	harps Square D 🦨 Coyote / T	NT Record
NTERNET			GCP6 🚡 UPLOAD 🔍	Search
KCHIVE	ABOUT BLOG PROJECTS HELP DONATI	CONTACT JOBS VOLUNT	EER PEOPLE	
lick on any	field below to edit it	CONTACT JOBS VOLUNTEER PEO	s Hara or O Salact files t	bbeo
IICK OIT arry		Mana And		o aud
age Title ^	Sets In Order 194901	Sets_in_Order_194811.pdf	5.8 MB	0
age URL *	https://archive.org/details/sets-in-order-194901	Sets_in_Order_194901.pdf	6.1 MB	0
escription *	Add a description of the item page			
ubject Tags	*Add keywords, separated by commas	Name Sets_In_Order_104811 pdt		
reator	Creator of the content	14 A		
ate	Date work was created/published	1		
ollection *	Community texts	~		
est item	No	~		
anguage	Language of the work	14 - C		
icense	No license selected	1		
	Add additional metadata			

Meta data window appears. Modify page title as needed in page title box. This is the title for the whole collection of issues so it should have magazine name and date range of issues.

🛈 🔒 🛛 https://arr	chive.org/upload/ C Q	ies to th	e internet archive 🔸 🐥 🏠 📋	•	* - 0
ost Visited 🧕 Wi	iting and Illustratin. 🜎 https://www.facebook 🔨 Diana Patton A	urtist	🗗 New Square Music - h. 🛞 Bee Sharps S	quare D. 🥳 Coyote / 1	NT Record
INTERNE ARCHIVE	' 🗆 🖩 🖯 🕄 🕼		CCP6		Search
	ABOUT BLOG PROJECTS HELP DONAT	ie 🛡	CONTACT JOBS VOLUNTEER	PEOPLE	
Click on any	field below to edit it		Drag and Drop More Files Her	e or 👩 Select files	to add
Page Title *	Sets in Order/Square Dancing	~	Name	Size	x
			Sets_in_Order_194811.pdf	5.8 MB	0
Page URL*	https://archive.org/details/		Sets_in_Order_194901.pdf	6.1 MB	0
	, in the second s				
Description *	Add a description of the item page	Contract Joe			
Subject Tags	*Add keywords, separated by commas	CONTACT JOBS			
Creator	Creator of the content				
Date	Date work was created/published	DONATE CONTACT JOBS VI			
Collection *	Community texts	~	-		
Test Item	No	~			
Language	Language of the work				
Linense	No monoto potentida				

Put unique id in URL box. It will become the Identifier for all the issues and become the next text in the created URLs. The last field of the URL will be the full file name.

Since the URL field had bobosgood, the first .PDF file's final full permalink URL will be: https://arvhive.org/download/bobosgood/Sets_in_Order194811.pdf .



Add the description by clicking on the line to open the box. Next add *Subject Tags*. Carefully consider what to put in these two fields since they will be particularly useful to people discovering items you've uploaded if they're doing a general search but don't already know the title of the magazine. Then put in the person or organization who created the magazines. Put in the date of the first magazine or leave blank. Leave the default collection since initially they offer you a limited selection. After your issues are loaded, however, contact an administrator at info@archive.org and request your item be added to the collection you think is best or a new collection you suggest. Select the appropriate language from the language box.

D 🔒 https://ar	chive.org/uplead/ C Q es	to the internet a	rchive -> 💺 🏫 🏠 📋		• \$
st Visited 🧕 W	friting and Illustratin. 👩 https://www.facebook 🦎 Diana Patton Artis	it <mark></mark> New S	quare Music - h 🥳 Bee Sharpa So	uare D_ 😴 Coyote / TN	T Record .
Page URL*	https://archive.org/details/bobosgood_202203	-	Sets_in_Order_194811.pdf Sets_in_Order_194901.pdf	5.8 MB 6.1 MB	0
Description '	Add a description of the item page	2			
Subject Tags	Add keywords, separated by commas	2			
Creator	Creator of the content	Sets in Order			
Date	Date work was created/published	* * *			
Collection *	Community texts				
Test Item	No	 × ×			
Language	Language of the work	2			
License	No license selected	1			
	C CC0 — "No Rights Reserved"				
	C Creative Commons				
	Require Share-Alike				
	Prohibit Commercial Use				
	C Public Domain				

Select a license. The Creative Commons, Attribution, NonCommercial, No derives seems to be the most protective. Be careful because you can give away your copyright rights here. Also make sure you have the rights to give away if the magazines are not your creation.

① A https://arc	hive.org/upload/	Q es to the	internet archive 🤿 🦊 🏫 🏠 自		er - 4
st Visited 🧕 Wr	nting and Illustratin 📢 https://www.facebook 🦎 Diana Pi	atton Artist	🥂 New Square Music - h 🛞 Bee Sharps Sq	uare D 🚽 🎻 Cayate / T	NT Record
Click on any	field below to edit it		Drag and Drop More Files Here	or OSelect files	to add
Page Title *	Sets In Order/Square Dancing	~	Name	Sce	x
Page URL *	https://archive.org/details/bobosgood_202203	~	Sets_in_Order_194811.pdf	5.8 MB 6.1 MB	0
Description *	Add a description of the item page	×.			
Subject Tags	*Add keywords, separated by commas	e.			
Creator	Creator of the content	×			
Date	Date work was created/published	1			
Collection *	Community texts	~			
Test item	No	1			
Language	Language of the work	~			
License	Creative Commons Attribution- NonCommercial-NoDerivs	~			
More Options	Add additional metadata				
	Upload	d and Create	Your Item		

Click the Upload and Create Your Item button.

Procedure to add files to an existing item

First set up your files so each .pdf file has a name similar to the ones already in the item. Go to the page for the item.



Click on Edit.



Upload PDF files to Internet Archive Click on *change the files*.

fisited 👩 Writin	g and Illustratin. 👩 https	//www.facebook 🔨 Dian	a Patton Artist 0 P	ew aquare music - n_	🖑 Bee Sharps Square D. 🥳	Coyote / TNT Record
TERNET	e II A	8 8 🖾			CCP6 T	UPLOAD Q See
CHIVE						-
	ALCON DOOS	allowed a street		3003	Hermoor Prove	
ilo od	litor for b	ohosaoo	4			
lie eo	litor for D	oposgood	L			
	And the second second					
dd a f	le 🔶					
		CORD-CALLER .				
ht-click / C1	RL-click on a file be	w lote it				
k on any ve	Second Street of Street Stre					
the state was a first of the state of the st	liow folders to expand	the sub-				
en en any je	illow folders to expand	the sub-				
en on any ye	ellow folders to expand	the sub-				
Tree Root	ellow folders to expand	the sub-				
Tree Root	ellow folders to expand	the sub-				
Tree Root	allow folders to expand	the sub-				
Tree Root	allow folders to expand a_archive torrent	the sub-				
Tree Roat	allow folders to expand	the subkcentre survey				
Tree Root	f_archive forrent f_archive forrent f_fice.ord f_meta.sqtte f_meta.sqtte	the sub-				
Tree Root bobosgoov bobosgoov bobosgoov bobosgoov bobosgoov bobosgoov bobosgoov bobosgoov	J_archive forrent 1_free.ori 5_meta.sqite 1_meta.sqite 1_meta.ml rder_194811.pdf	the sub-				
Tree Root Tree Root bobosgoor	Allow folders to expand	the such a such as a				
Tree Root bobosgoov bobosgoov bobosgoov bobosgoov Sets_in_0 sets_in_0	4, archive forrent 4_fractive forrent 4_fies.cml 5_mets sqlte 5_mets sqlte 1_mets	the such and				
Tree Root bobosgoov bobosgoov bobosgoov bobosgoov Sets_In_0 sets_In_0 Sets_In_0 Sets_In_0	4_archive forment 1_files.ord 1_mets.sqtbe 1_mets.sqtbe 1_mets.tml rder_104811_chacr.bitml.gz rder_104801_pdf rder_104801_pdf	the such and				
Tree Root E bobosgoor bobosgoor bobosgoor bobosgoor bobosgoor Sets_in_0 Sets_in_0 Sets_in_0 Sets_in_0	Larchive forcent f_archive forcent f_meta.snlte meta.	the sub-				
Tree Roat Dobosgoo bobos	4_archive fortent _files.or/ _files.or/ f_meta.qtbe _meta.wti rder_194811.pdf rder_194811.pdf rder_194911.pdf rder_194901.pdf rder_194901.pdf rder_194901.pdf rder_194901.pdf	the such and the				

Click on Add a File.

sited D Writing and Bustratin () https://www.facebook	Y Diana Patton Arbst 💕 New Square Music - h 🧐 Bee	/ Sharps Square D 🥳 Cayote / TNT Reco	urd
Square_Dancing_198301 pdf	16 1	лв 😒	
Square_Dancing_198302.pdf	20 M	AB 🙁	
Square_Dancing_198303 pdf	14 N	AB 🙁	
Square_Dancing_198304 pdf	16.1	IB 🙁	
Square_Dancing_198305 pdf	14 h	лв 😒	
Square_Dancing_198306.pdf	16 1	AB 😣	
Square_Dancing_198307.pdf	15 M	AB 😣	
Square_Dancing_198308.pdf	14 5	IB 🙁	
Square_Dancing_198309 pdf	18 1	лв 🙁	
Square_Dancing_198310 pdf	21 h	AB 🙁	
Square_Dancing_198311.pdf	31 M	1B 🙁	
Gquare_Dancing_198312.pdf	16 N	IB 😣	
Square_Dancing_198401.pdf	16 M	AB 😢	
Square_Dancing_198402 pdf	20 M	AB 🙁	
Square_Dancing_198403.pdf	17.6	AB 🙁	
Square_Dancing_198404.pdf	14 6	AB 😒	
Square_Dancing_198405.pdf	16.1	IB 🙁	
Square_Dancing_198406.pdf	16 1	лв 😒	
Square_Dancing_198407.pdf	151	AB 😣	
Square_Dancing_198410.pdf	16.1	IB 😣	
Square_Dancing_198411.pdf	17.6	IB 🙁	
Square_Dancing_198412 pdf	221	лв 👩	

Drag over the new files to add. Click Add Files To Existing Item.

TNT Record.	uare D., 🦿 Coyote / Ti	🗗 New Square Music - h 🛞 Bee Sharps Square I	https://www.facebook 🦎 Dians Patton Artist	Visited 🛄 Writing and Illustratin 📢
	0	0.0100		
		O.D.MD		Sets_In_Order_196707 pdf
	0	13 NB		Bets_In_Order_196708.pdf
	0	IT MB		Bets_in_Order_196709 pdf
	0	15 MB		Sets_In_Order_196710.pdf
	0	15 MB		Sets_In_Order_196711 pdf
	0	13 MB		Sets_in_Order_196712.pdf
	0	8.7 MB		Sets_in_Order_196801.pdf
	0	11-MB		Sets_in_Order_196802.pdf
	0	10 MB		Sets_in_Order_196803.pdf
	0	12 MB		Sets_In_Order_196804 pdf
	0	11 MB		Sets_in_Order_196805.pdf
	0			Sets_in_Order_196806.pdf
	0	page is being created	Please wait while your	Sets_in_Order_196807 pdf
	0			Sets_In_Order_196808 pdf
	0			Sets_in_Order_196809 pdf
	0	100 100 00		Sets_In_Order_196810.pdf
	0	601.pdf 109 MB/2.0 GB	Sets_In_Order_195	Sets_In_Order_196811 pdf
	0			Sets_in_Order_196812 pdf
	0	11 MB		Bets_in_Order_196901 pdf
	٥	15 MB		Sets_III_Order_196902.pdf
	0	12 MB		Sets_in_Order_196903 pdf
				Cote in Order 106004 edf
	000000000000000000000000000000000000000	601.pdf 109 MB/2.0 GB 11 MB 15 MS 12 MB	Please wait while your	Sets In_Order_146807.pdf Sets In_Order_146806.pdf Sets In_Order_146806.pdf Sets In_Order_146806.pdf Sets In_Order_146806.pdf Sets In_Order_146812.pdf Sets In_Order_146802.pdf Sets In_Order_146803.pdf Sets In_Order_146803.pdf

Wait for files to be uploaded.

Once the files are uploaded, you can create your own index to the pages using the pdf files shown if you click the PDF file





In picture above a list with the .PDF file names is shown. These names are pasted on to the standard Internet Archive download URL to create the link for your index. The full link would be https://archive.org/download/bobosgood/Sets_In_Order_194811.pdf. The full link can be copied to the clipboard with ctrl-C while hovering over the link.

Internet Archive Derivative Files for PDF's

Once you have uploaded your .PDF files, the Internet Archive will create a number of derivative files. They do this because the files they are given could have been created by a variety of different software packages which may have provided different kinds of data. For example they will create individual .JPG images of teach page and do Optical Character Recognition (OCR) to pull off any text that might have been in the .JPG images and they will pull out pictures on a page and compress them for efficient transmission over the internet. In one case, I uploaded 118 files of the Magazine Squares and Rounds and after the derivative files were created there were 1185 files.

The derivative files are created by a background process which may take hours or days to process a group of input files. When I uploaded 444 issues of a magazine it took about 2 weeks to create the derivative files.

After the derivative files are created, when you view your item, there will be an index pointer to each issue in your item. The pointers are shown in the Internet Archive viewer window. When you click on an issue in the index, that issue will be shown in the viewer.



If you don't see the index it means that the derivative files are still being created or that there is a problem in their creation. You can click on the history button (see diagram below) to see what processing has been done.



In my case, my the .PDF files I uploaded had been created using PDF Converter Pro and had a text layer created by that program's OCR software. It had also been processed with MRCS so the images had been extracted and separated from the text. Then the files had been optimized for fast internet download. Thus the Internet Archive process duplicated most of what I had done to get my files into their archive required format.

In one Internet Archive post, I found that the Internet Archive does not modify the original .PDF files. If there is a text layer in the .PDF file then the file is searchable. If there is no text layer then they create a derivative file with their OCR to make the file searchable when viewed with their viewer. This would indicate to me that if you are going to pointing to the .PDF file which you uploaded, and your index link will cause it to be downloaded, it should already have been made searchable before it is uploaded. Note, text searching is only as good as the OCR. If the OCR does not read the characters correctly, a search for words you may be able to recognize eye might not may not be found if the OCR recognition is bad.

So the question is, if you create an index to the files you saved in the Archive what link should the index point to?

If you have an issue, for example Squares and Rounds for November 1976 and you point to <u>https://archive.org/details/squares-rounds/Squares_and_Rounds_197611</u> the archive will pop up their viewer with that issue showing in that viewer where you can page through, or search for text in the issue. Your search will use their derived files.

(note: the action keyword in the URL is download not details and there is a .PDF suffix is at the end) you will get your original .PDF file downloaded with whatever features (search, MRCS, etc.) you had included in it when it was uploaded.

Constraints

As of June 4, 2021 one Internet Archive page said:

Currently, there is no limit on the size of files nor the number of files. However, from a systems perspective, we do not recommend files larger than 50 GBs to be uploaded or more than 1000 files, per single page.

Gcp 29 March 2022