

---

# **How to make PDF from Markdown with Pandoc**

Detailed manual for all

Author: Alexey Gumirov



11 January 2021

## Contents

<b>1</b>	<b>How to make PDF from MarkDown with Pandoc</b>	<b>5</b>
1.1	How-to for docs preparation . . . . .	5
1.1.1	Tools . . . . .	5
1.1.2	Instructions and commands . . . . .	6
1.1.3	Pandoc command . . . . .	7
1.1.4	Important notes about MarkDown file formatting for PDF processing . . . . .	12
1.2	Protection of PDF file with QPDF . . . . .	13
1.3	Examples . . . . .	13
	<b>Example of the unnumbered chapter</b>	<b>15</b>
1.4	Unnumbered header of level 4 and lower . . . . .	15
<b>2</b>	<b>Automation of PDF creation</b>	<b>16</b>
2.1	Local PC automation with <i>entr</i> and <i>task spooler</i> . . . . .	16
2.2	Building CI pipeline in the Gitlab . . . . .	16
2.2.1	Folders structure . . . . .	16
2.2.2	Single stage pipeline . . . . .	17
2.2.3	Pipeline to produce protected PDF . . . . .	18

## List of Tables

1	Sample table . . . . .	11
---	------------------------	----

## List of Figures

1	Aleph 0 . . . . .	10
---	-------------------	----

# 1 How to make PDF from MarkDown with Pandoc

How-To, templates and commands to produce PDF documents from MarkDown files.

## 1.1 How-to for docs preparation

### 1.1.1 Tools

- **Pandoc**

- Template: I use my template which is a slightly modified eisvogel.latex<sup>1</sup> template. I made following changes:
  - \* Each paragraph starts from the new page.
  - \* Block quote font is darker than original which is better for reading.
    - Original color is `rgb{119,119,119}`, mine setting is `rgb{89,89,89}`.
  - \* Code listings are wrapped on white spaces by default.
  - \* Code listings font size is set to “footnotesize”. And original template parameter does not work (it, actually, never worked properly).
  - \* Line interval in code listings is set to 1.2.
- Both templates you can find in the repository of this project. Original template eisvogel.latex<sup>2</sup> and my modified eisvogel\_mod.latex<sup>3</sup>

- **texlive**

- **convert**

- converts and formats images.
- it is used here for the change of DPI of the images and convert to PNG.
- **convert** is the utility which is part of the **ImageMagick** package.

I did not install **convert** tool, it seems like it is installed by default in Ubuntu or comes with **texlive**. To avoid possible issues with **pdflatex** engine I did full installation of **texlive** packet.

In Debian family (with **apt**):

```
sudo apt-get update
sudo apt-get install pandoc
sudo apt-get install imagemagick
```

Quite often standard Debian and Ubuntu repositories install very old version of Pandoc (something like 1.19), which does not support smart extensions and many other features. Then it is better to download fresh [deb](#) package from the github repository: Pandoc Github<sup>4</sup>. Installation of the [deb](#) package is made with the following command:

```
dpkg -i <package name>.deb
```

<sup>1</sup><https://github.com/Wandmalfarbe/pandoc-latex-template>

<sup>2</sup>[pandoc/templates/eisvogel.latex](#)

<sup>3</sup>[pandoc/templates/eisvogel\\_mod.latex](#)

<sup>4</sup><https://github.com/jgm/pandoc/releases>

I use following `texlive` packages:

```
sudo apt-get install texlive-latex-recommended
sudo apt-get install texlive-fonts-recommended
sudo apt-get install texlive-latex-extra
sudo apt-get install texlive-fonts-extra
sudo apt-get install texlive-xetex
```

Extra LaTeX packages are needed for **eisvogel** template to work. I also install XeTeX because if you have text with some special symbols, XeTeX can process it properly.

### 1.1.2 Instructions and commands

**1.1.2.1 YAML Block for LaTeX template** This YAML block in the beginning of the MarkDown file defines parameters used by the Pandoc engine and relevant LaTeX template parameters. This particular example below instructs Pandoc to produce PDF file with the Cover page (**titlepage: true**) and change color of the line on the cover page. Another important parameter is **logo** - it defines path to file with the logo you want to put on the cover page.

```
title: "How to make PDF from MarkDown with Pandoc"
author: "Author: Alexey Gumirov"
date:
subtitle: "Detailed manual for all"
geometry: "left=2.54cm,right=2.54cm,top=1.91cm,bottom=1.91cm"
titlepage: true
titlepage-color: "FFFFFF"
titlepage-text-color: "000000"
titlepage-rule-color: "CCCCCC"
titlepage-rule-height: 4
logo: "files/logo.png"
logo-width: 100
page-background:
page-background-opacity:
links-as-notes: true
lot: true
lof: true
listings-disable-line-numbers: true
listings-no-page-break: false
disable-header-and-footer: false
header-left:
header-center:
header-right:
footer-left: "© Alexey Gumirov"
footer-center: "License: WTFPL"
footer-right:
subparagraph: true
lang: en-US
```

Table of content, list of tables and list of figures are going in the following order: ToC, LoT and LoF. Each pages starts from the new line.

Parameter **links-as-notes** enables putting of the URL links in the footnotes of the page.

Parameters **lof** and **lot** are responsible for the creation of *list of figures* and *list of tables* respectively.

Parameter **listings-disable-line-numbers** disables line numbers for all listings.

Because MarkDown for GitHub does not support YAML header in the main file, I set it up in the separate **HEADER.YAML** file in the root folder of the project.

**1.1.2.2 Images preparation** In my setup I print with 300 DPI (this produces high resolution PDF). Therefore all images must be 300 DPI. If you have images with different DPI (especially GIF files), then use the following commands:

To re-sample image to 300 DPI:

```
convert $SOURCE_IMG_FILE -units PixelsPerInch -resample 300
$TARGET_IMG_FILE.png
```

After re-sampling image has to be brought to the proper size. Command resizes picture to 1700 pixels of width and sets DPI meta-data to 300.

```
convert $SOURCE_IMG_FILE -units PixelsPerInch -resize 1700x -density 300
$TARGET_IMG_FILE.png
```

But if you are not afraid, then all can be done in one command:

```
convert $SOURCE_IMG_FILE -set units PixelsPerInch -resample 300 -resize
1700x -density 300 $TARGET_IMG_FILE.png
```

It is important to mention that the order of options does matter. The instruction above makes steps in the following order:

1. `-set units PixelsPerInch`: Sets density units in Pixels per Inch instead of default `PixelsperCentimeter`.
2. `-resample 300`: Changes resolution of the image from its current DPI (PPI) to 300 DPI (PPI). It is not just change of meta-data, this parameter makes **convert** to re-process image.
3. `-resize 1700x`: Resizes picture to the following dimensions: `width = 1700 pixels`, `height = auto`.
4. `-density 300`: This parameter sets DPI meta-data in the target picture to 300 DPI (PPI)

### 1.1.3 Pandoc command

Putting all together in one command.

All Pandoc commands are for the Pandoc version 2.x.

Since version 2.11 Pandoc warns that source format `markdown_github` is deprecated. For my formatting following replacement works: `markdown_github->markdown_strict+pipe_tables+backtick_code_blocks`. Below all scripts are given with the new `markdown_strict` source format.

```
pandoc -s -o $DEST.pdf -f "markdown_strict+pipe_tables+
backtick_code_blocks+auto_identifiers+yaml_metadata_block+
implicit_figures+table_captions+footnotes+smart+header_attributes" \
--template eisvogel_mod --listings --columns=50 --number-sections \
--toc --dpi=300 HEADER.YAML $SOURCE.md
```

Because I use YAML header, all `-V` parameters I put there.

If you want to put current date in the cover page automatically, then you can add following parameter in the **pandoc** command line:

```
-M date="`date "+%d %B %Y"`"
```

Or you can define date in the script variable `DATE=$(date "+%d %B %Y")` and then use this variable in the `-M` option: `-M date="$DATE"`.

Then **pandoc** command will look like that:

```
DATE=$(date "+%d %B %Y")
pandoc -s -o $DEST.pdf -f "markdown_strict+pipe_tables+
backtick_code_blocks+auto_identifiers+yaml_metadata_block+
implicit_figures+table_captions+footnotes+
smart+header_attributes" --template eisvogel_mod --listings --columns=50
--number-sections --toc --dpi=300 -M date="$DATE" HEADER.YAML
$SOURCE.md
```

Options of the **pandoc** command mean following:

- `-s`: Standalone document.
  - Produce typographically correct output, converting straight quotes to curly quotes, — to em-dashes, – to en-dashes, and ... to ellipses. Nonbreaking spaces are inserted after certain abbreviations, such as “Mr.” (Note: This option is selected automatically when the output format is latex or context, unless `--no-tex-ligatures` is used. It has no effect for latex input.)
    - \* In newer versions of **pandoc** this switch was removed and you shall use `+smart` extension in the `-f` option.
- `-f FORMAT` or `-r FORMAT`:
  - Specify input format. `FORMAT` can be **native** (native Haskell), `json` (JSON version of native AST), `markdown` (pandoc’s extended Markdown), `markdown_strict` (original unextended Markdown), `markdown_phpextra` (PHP Markdown Extra), `markdown_github` (GitHub-Flavored Markdown), `commonmark` (CommonMark Markdown), `textile` (Textile), `rst` (reStructuredText), `html` (HTML), `docbook` (DocBook), `t2t` (txt2tags), `docx`



(docx), `odt` (ODT), `epub` (EPUB), `opml` (OPML), `org` (Emacs Org mode), `mediawiki` (MediaWiki markup), `twiki` (TWiki markup), `haddock` (Haddock markup), or `latex` (LaTeX). If `+lhs` is appended to `markdown`, `rst`, `latex`, or `html`, the input will be treated as literate Haskell source. Markdown syntax extensions can be individually enabled or disabled by appending `+EXTENSION` or `-EXTENSION` to the format name. So, for example, `markdown_strict+footnotes+definition_lists` is strict Markdown with footnotes and definition lists enabled, and `markdown-pipe_tables+hard_line_breaks` is pandoc's Markdown without pipe tables and with hard line breaks.

- `auto_identifiers`: A heading without an explicitly specified identifier will be automatically assigned a unique identifier based on the heading text. Allows to make cross references. More information on Pandoc documentation page<sup>5</sup>.
  - `backtick_code_blocks`: In addition to standard indented code blocks, pandoc supports fenced code blocks. These begin with a row of three or more backticks (‘) and end with a row of backticks that must be at least as long as the starting row.
  - `footnotes`: Footnotes in the Pandoc Markdown format. For more details please go to Pandoc manual page<sup>6</sup>.
  - `header_attributes`: Headings can be assigned attributes using this syntax at the end of the line containing the heading text: `{#identifier .class .class key=value key=value}`. For example, to make chapter unnumbered use `{.unnumbered}` or `{-}`.
  - `implicit_figures`: An image with nonempty alt text, occurring by itself in a paragraph, will be rendered as a figure with a caption. The image's alt text will be used as the caption. This extension is very useful when you need to autogenerate captions for figures in the markdown reference format like: `![This is the caption](/url/of/image.png)`
  - `pipe_tables`: Table syntax identical to Github and PHP Markdown Extra Tables.
  - `smart`: Produce typographically correct output, converting straight quotes to curly quotes, `—` to em-dashes, `-` to en-dashes, and `...` to ellipses. Nonbreaking spaces are inserted after certain abbreviations, such as “Mr.” (Note: This option is selected automatically when the output format is latex or context, unless `--no-tex-ligatures` is used. It has no effect for latex input.)
  - `strikeout`: To strikeout a section of text with a horizontal line, begin and end it with `~~`.
  - `table_captions`: A caption may optionally be provided for all 4 kinds of supported Markdown tables. A caption is a paragraph beginning with the string `Table:` (or just `:`), which will be stripped off. It may appear either before or after the table.
  - `yaml_metadata_block`: A YAML metadata block is a valid YAML object, delimited by a line of three hyphens (`---`) at the top and a line of three hyphens (`---`) or three dots (`...`) at the bottom. A YAML metadata block may occur anywhere in the document, but if it is not at the beginning, it must be preceded by a blank line.
- `--template FILE`: Use `FILE` as a custom template for the generated document. Implies `--standalone`.
  - `--toc: --table-of-contents`
    - Include an automatically generated table of contents (or, in the case of latex, context, docx, and rst, an instruction to create one) in the output document. This option has no effect

<sup>5</sup>[https://pandoc.org/MANUAL.html#extension-auto\\_identifiers](https://pandoc.org/MANUAL.html#extension-auto_identifiers)

<sup>6</sup><https://pandoc.org/MANUAL.html#footnotes>

on man, docbook, docbook5, slidy, slideous, s5, or odt output.

- `--dpi:`
  - Specify the dpi (dots per inch) value for conversion from pixels to inch/centimeters and vice versa. The **default** is **96dpi**. Technically, the correct term would be ppi (pixels per inch).
- `-V KEY[=VAL]: --variable=KEY[:VAL]`
  - Set the template variable KEY to the value VAL when rendering the document in standalone mode. This is generally only useful when the `--template` option is used to specify a custom template, since pandoc automatically sets the variables used in the default templates. If no VAL is specified, the key will be given the value true.
  - `lang`: one of the KEY parameters of `-V` which defines default document language. Changing of this parameter will change language of default headers and captions (e.g. if you make `lang=de-DE`, then **Contents** will become **Inhaltsverzeichnis**, **List of Tables** will be **Tabellenverzeichnis**, **Table** will be **Tabelle**, **Figure** caption will be **Abbildung**).
  - `subparagraph`: Is needed to start each chapter from the new page here. In the Eisvogel\_mod.latex template necessary modifications are made.

Additional useful options of the **pandoc** command are:

- `--listings`: It creates nice presentation of the raw code (like shell code or programming code).
- `--columns`: Specify length of lines in characters. This affects text wrapping in the generated source code (see `-wrap`). It also affects calculation of column widths for plain text tables.
- `--number-sections`: Automatically creates enumerated headers.
- `--default-image-extension`: If you want Pandoc to insert only one type of images, e.g. PNG, then you shall add `--default-image-extension png` in the command line.

**1.1.3.1 List of figures** List of figures is automatically generated by the Pandoc during PDF file creation. For the list of figures and relevant captions is responsible `implicit_figures` extension. It does not require any additional text, it will convert [alt text] into the caption. E.g. for this image below:

```
![Aleph 0](files/logo.png)
```



**Figure 1:** Aleph 0

**1.1.3.2 List of tables** The `table_captions` extension requires `Table:` or `:` paragraph right before or below table. You do not need to numerate the table - Pandoc will make enumeration by itself,

but you shall provide required paragraph text. E.g. for the table below the raw Markdown text is the following:

Table: Sample table

```
| Name | value |
|:-----|:-----:|
| A    | 1     |
| B    | 2     |
```

**Table 1:** Sample table

Name	value
A	1
B	2

For the convenient formatting of your tables in Markdown files, I recommend to use the following VIM plugin: *VIM Table Mode*<sup>a</sup>.

<sup>a</sup><https://github.com/dhruvasagar/vim-table-mode>

**Processing of multiple files** When you create large amount of content, it is not convenient to use one large MarkDown file for it. Then it is better to split it in multiple MarkDown files and organize them in a separate folder using names with leading sequence numbers, like here:

- Create folder, e.g. **“content”**.
- Put there Markdown files which you want to combine into one PDF.
- Name files with numbers in the order they shall be concatenated into one PDF. Example:

```
> ~/ $ ls -lh content/
total 197K
-rwxrwxrwx 1 root root 0 Dec 18 18:49 00-Intro.md
-rwxrwxrwx 1 root root 0 Dec 18 18:47 01-Chapter_A.md
-rwxrwxrwx 1 root root 0 Dec 18 18:47 02-Chapter_B.md
-rwxrwxrwx 1 root root 0 Dec 18 18:49 03-Chapter_C.md
-rwxrwxrwx 1 root root 0 Dec 18 18:50 99-Appendix.md
```

- Apply following Pandoc command:

```
pandoc -s -o $DEST.pdf -f "markdown_strict+pipe_tables+
backtick_code_blocks+auto_identifiers+yaml_metadata_block+
implicit_figures+table_captions+footnotes+
smart+header_attributes" --template eisvogel_mod --listings --columns=50
--number-sections --toc --dpi=300 HEADER.YAML content/*.md
```

This command will take all MarkDown files from the **“content”** folder and convert them into enumerated order into a single PDF file.

The cons of this method is that you cannot include/exclude particular source MarkDown files to produce PDF with only content you need. Therefore for such setups I use **INDEX** file where I list all files which Pandoc shall convert into PDF in the order I want them to go.

```
> cat INDEX
HEADER.YAML
00-Intro.md
01-Chapter_A.md
03-Chapter_C.md
```

And then my PDF generation command looks the following:

```
pandoc -s -o $DEST.pdf -f "markdown_strict+pipe_tables+\
backtick_code_blocks+auto_identifiers+yaml_metadata_block+\
implicit_figures+table_captions+footnotes+\
smart+header_attributes" --template eisvogel_mod --listings --columns=50
--number-sections --toc --dpi=300 $(cat INDEX)
```

#### 1.1.4 Important notes about MarkDown file formatting for PDF processing

**1.1.4.1 Unordered Lists and sub-lists indentation** It is stated in the GitHub<sup>7</sup> site that correct indent for the unordered lists is 2 spaces. But with this indent Pandoc does not identify sub-lists.

Therefore, please use 4 spaces indent for the sub-lists in the unordered lists. Then they will be properly reflected in the PDF files.

While using of standard tab (4 spaces) indent is not a mistake, some programs (in my case it is MS Visual Studio Code) can give you a warning. You can just ignore it.

**1.1.4.2 Links** If your Markdown file has to be processed into the PDF, then please pay attention to the format of links you use:

a) Link format that does NOT WORK: **![Name of the resource](Link)**.

b) Link format that WORKS: **[Name of the resource](Link)**.

The problem is that by the Markdown guidelines<sup>8</sup> using exclamation mark before URL is not appropriate. Exclamation mark is used for links to images only. But GitHub engine does not give you an error, it just treats such links as links which opens in the new tab or window in the browser. Therefore, to avoid compilation errors in the **pdflatex** engine (which is used by **pandoc**), please use (b) type of URL formatting, which is compliant with Markdown standard.

**1.1.4.3 Pandoc execution folder** For the correct processing of the links and references by Pandoc (especially links to images) you shall run pandoc script inside the directory with MarkDown files. Therefore, it is better to place **logo** folder, YAML meta-data file and PDF generating shell script directly into the directory with MarkDown files.

<sup>7</sup><https://github.com/DavidAnson/markdownlint/blob/v0.11.0/doc/Rules.md#md007>

<sup>8</sup><https://github.com/DavidAnson/markdownlint/blob/v0.11.0/doc/Rules.md#md007>

## 1.2 Protection of PDF file with QPDF

Pandoc does not produce password protected PDF files. To create password protected PDF and also being able to disable ability to extract data from the document and print it I use **qpdf**<sup>9</sup> command line tool.

```
qpdf --object-streams=disable --encrypt "{user-password}"
    "{owner-password}" 256 --print=none --modify=none --extract=n --
    {input.pdf} {output.pdf}
```

- **user-password:** This is the password to open the PDF file in the reader program.
- **owner-password:** This is the password which protects PDF from editing.

Usually I use only **owner-password** because I want my files be protected from editing.

It is important to mention that if you want to have no **user-password** while have **owner-password**, you shall define empty user password:

```
qpdf --object-streams=disable --encrypt "" "{owner-password}" 256
    --print=none --modify=none --extract=n -- {input.pdf} {output.pdf}
```

In order to generate random **owner-password** you can use many methods defined on this page “10 Ways to Generate a Random Password from the Linux Command Line”<sup>10</sup>.

For unification of PC and GitLab CI pipeline scripts I use the last one (see below), because it works in the **alpine** Docker container:

```
date | md5sum | cut -d ' ' -f 1
```

Finally, merging all into one script:

```
OWNER_PASSWORD=$(date | md5sum | cut -d ' ' -f 1)

qpdf --object-streams=disable --encrypt "" "$OWNER_PASSWORD" 256
    --print=none --modify=none --extract=n -- {input.pdf} {output.pdf}
```

## 1.3 Examples

This page *pandoc-2-pdf-how-to.pdf*<sup>11</sup> as normal PDF and also this page as protected PDF *\*pandoc-2-pdf-how-to\_(protected).pdf*<sup>12</sup> were generated by the following shell script<sup>13</sup>:

```
#!/bin/sh

SOURCE_FILE_NAME="README"
DEST_FILE_NAME="pandoc-2-pdf-how-to.pdf"
DEST_FILE_NAME_PROTECTED="pandoc-2-pdf-how-to_(protected).pdf"
INDEX_FILE="INDEX"
```

<sup>9</sup><http://qpdf.sourceforge.net/>

<sup>10</sup><https://www.howtogeek.com/howto/30184/10-ways-to-generate-a-random-password-from-the-command-line/>

<sup>11</sup>*pandoc-2-pdf-how-to.pdf*

<sup>12</sup>*pandoc-2-pdf-how-to\_(protected).pdf*

<sup>13</sup>*\_pdf-gen.sh*

```
TEMPLATE="eisvogel_mod.latex"
DATE=$(date "+%d %B %Y")
DATA_DIR="pandoc"

SOURCE_FORMAT="markdown_strict
+pipe_tables\
+backtick_code_blocks\
+auto_identifiers\
+yaml_metadata_block\
+implicit_figures\
+table_captions\
+footnotes\
+smart\
+header_attributes"

pandoc -s -o "$DEST_FILE_NAME" -f "$SOURCE_FORMAT" --data-dir="$DATA_DIR"
--template "$TEMPLATE" --toc --listings --columns=50 --number-sections
--dpi=300 --pdf-engine xelatex -M date="$DATE" $(cat "$INDEX_FILE") >&1

OWNER_PASSWORD=$(date | md5sum | cut -d ' ' -f 1)

qpdf --object-streams=disable --encrypt "" "$OWNER_PASSWORD" 256
--print=none --modify=none --extract=n -- "$DEST_FILE_NAME"
"$DEST_FILE_NAME_PROTECTED"
```

Links to [HEADER.YAML](#)<sup>14</sup> and [INDEX](#)<sup>15</sup> files.

---

<sup>14</sup>HEADER.YAML

<sup>15</sup>INDEX

## Example of the unnumbered chapter

If you want some chapters be without numbers (e.g. Annex or Preface), you can use so called **header attributes**, represented as a `{#identifier .class .class key=value key=value}` after the header.

For example, to exclude this chapter header from numbering, Markdown code can look like:

```
# Example of the unnumbered chapter {.unnumbered}
```

or

```
# Example of the unnumbered chapter {-}
```

**Note:** This syntax is not compatible with the Github flavored Markdown.

### 1.4 Unnumbered header of level 4 and lower

Sometimes there is a need to make an unnumbered header of level 4 and lower. The problem of Pandoc is that it puts line break after unnumbered header of levels 1 to 3, but when header is of level 4 and below, then line break is not added.

And we end up with this ugly situation.

**Level 4 unnumbered header** Text of chapter below header, but Pandoc does not put any break after it. Let's fix this broken formatting.

#### Level 4 unnumbered header with proper line break

In order to force Pandoc process such formatting correctly, you can use following trick:

1. Enable `escaped_line_breaks` extension. (Just add `+escaped_line_breaks` in the source format extensions parameter of the script.)
2. Put forced line break after the header. Like this:

```
#### Level 4 unnumbered header with proper line break {-}  
\nIn order to force Pandoc process such formatting correctly, you can use  
following trick:
```

## 2 Automation of PDF creation

### 2.1 Local PC automation with *entr* and *task spooler*

On my local PC I use *entr* and *task spooler* (in Ubuntu it is called *tsp*).

- *entr*: The *Event Notify Test Runner* is a general purpose Unix utility intended to make rapid feedback and automated testing natural and completely ordinary. More details on the Entr project page<sup>16</sup>.
- *task-spooler* or *tsp* or *ts* (depending on the system): A simple unix batch system. More details via `man tsp` or `man ts`.

To install *entr* and *task spooler* in Ubuntu, use these commands:

```
sudo apt-get update
sudo apt-get install entr
sudo apt-get install task-spooler
```

The following command creates task in the spooler queue which monitors state of the edited file (in this case `README.md`) and as soon as file is updated, script `_pdf-gen.sh` is launched. This script generates PDF. In this example both `README.md` and `_pdf-gen.sh` are located in the same directory, and command below is launched from the same directory.

```
> $ tsp bash -c 'ls README.md | entr -p ./_pdf-gen.sh'
```

When you need to monitor multiple MarkDown files in the e.g. `content` folder, you can use the following command:

```
> $ tsp bash -c 'ls content/*.md | entr -p ./_pdf-gen.sh'
```

### 2.2 Building CI pipeline in the Gitlab

I made my CI pipeline for GitLab which automatically creates PDF and stores it in the Gitlab artifactory when the content of MarkDown or YAML files is changed.

#### 2.2.1 Folders structure

Create following folders structure:

```
> $ tree -a
./
-- content/
    -- 01-Introduction.md
    -- 02-Chapter_A.md
    -- 03-Chapter_B.md
    -- {...}.md
    -- HEADER.YAML
    -- INDEX
```

---

<sup>16</sup><http://eradman.com/entrproject/>



```

-- img/
  -- img_01.png
  -- img_02.png
  -- img_03.png
-- logo/
  -- logo.png
-- pandoc/
  -- templates/
    -- eisvogel.latex
    -- eisvogel_mod.latex
-- .gitlab-ci.yml
-- README.md

```

Where `INDEX` file contains list of source files which shall be processed by Pandoc including `HEADER.YAML` file.

```

> $ cat INDEX
HEADER.YAML
01-Introduction.md
02-Chapter_A.md
03-Chapter_B.md
{...}.md

```

- In `logo` folder I put `logo.png` file.
- In the `content` folder I create `img` folder where I put all images/pictures I use in the content MarkDown files.
- In the `content/pandoc/templates` folder I keep LaTeX templates I use for PDF creation.

To create PDF I use `knsit/pandoc` Docker container. This container has newer version of the **pandoc** therefore instead of `-S` optoin I use `+smart` extension in the `-f` option.

### 2.2.2 Single stage pipeline

The example of the pipeline below will allow you to produce PDF automatically using GitLab CI engine.

The `.gitlab-ci.yml` has the following content:

```

image: knsit/pandoc:latest
my_nice_pdf:
  variables:
    SOURCE_DIR: "content"
    INDEX_FILE: "INDEX"
    DEST_FILE_NAME: "my_nice_document"
    TEMPLATE: "eisvogel_mod"
    SOURCE_FORMAT: "markdown_strict+pipe_tables+\
      backtick_code_blocks+auto_identifiers+\
      yaml_metadata_block+smart+implicit_figures+\
      table_captions+footnotes+smart+header_attributes"
    DATA_DIR: "pandoc"
  script:

```

```

- DATE=$(date +%Y-%m-%d)
- DEST_FILE_NAME_DATE=$DEST_FILE_NAME$DATE
- DATE=$(date "+%d %B %Y")
- cd "$SOURCE_DIR"
- pandoc -s -o $DEST_FILE_NAME_DATE.pdf -f $SOURCE_FORMAT
  --data-dir="$DATA_DIR" --template $TEMPLATE -M date="$DATE"
  --listings --columns=50 --number-sections --toc --dpi=300 $(cat
    "$INDEX_FILE") >&1
- mkdir -p my_nice_pdf
- mv $DEST_FILE_NAME_DATE.pdf "$CI_PROJECT_DIR"/my_nice_pdf/
stage: build
artifacts:
  paths:
  - my_nice_pdf
  expire_in: 6 month
only:
  changes:
  - content/HEADER.YAML
  - content/INDEX
  - content/content.md

```

Parameter `changes` makes CI job run only when content of the YAML block or any of MarkDown files in the `content` folder is changed.

### 2.2.3 Pipeline to produce protected PDF

The example of the pipeline below uses two stages to produce PDF protected from editing and copying: - First stage - to produce PDF using **knsit/pandoc** container. - Second stage - to produce protected PDF using **alpine** container.

The `.gitlab-ci.yml` has the following content:

```

stages:
- makepdf
- protect

make_unprotected:
  image: knsit/pandoc:latest
  variables:
    SOURCE_DIR: "content"
    INDEX_FILE: "INDEX"
    DEST_FILE_NAME: "content.pdf"
    TEMPLATE: "eisvogel_mod"
    SOURCE_FORMAT: "markdown_strict+pipe_tables+\
      backtick_code_blocks+auto_identifiers+\
      yaml_metadata_block+smart+implicit_figures+\
      table_captions+footnotes+smart+header_attributes"
    DATA_DIR: "pandoc"
  stage: makepdf
  script:
    - DATE=$(date "+%d %B %Y")

```

```
- cd "$SOURCE_DIR"
- pandoc -s -o "$DEST_FILE_NAME" -f $SOURCE_FORMAT
  --data-dir="$DATA_DIR" --template $TEMPLATE -M date="$DATE"
  --listings --columns=50 --number-sections --toc --dpi=300 $(cat
  "$INDEX_FILE") >&1
- mkdir -p interim/
- mv "$DEST_FILE_NAME" "$CI_PROJECT_DIR"/interim/

artifacts:
  paths:
    - interim
  expire_in: 30 min
only:
  changes:
    - content/HEADER.YAML
    - content/INDEX
    - content/content.md

make_protected:
  image: alpine:latest
  variables:
    DEST_FILE_NAME: "my_nice_pdf"
    SOURCE_PDF_FILE: "content.pdf"
  stage: protect
  when: on_success
  script:
    - DATE=$(date +%Y-%m-%d)
    - DEST_FILE_NAME_DATE=$DEST_FILE_NAME$DATE".pdf"
    - apk add --update qpdf
    - PASSWORD=$(date | md5sum | cut -d ' ' -f1)
    - qpdf --object-streams=disable --encrypt "" "$PASSWORD" 256
      --print=none --modify=none --extract=n --
      interim/"$SOURCE_PDF_FILE" "$DEST_FILE_NAME_DATE"
    - mkdir -p my_nice_pdf/
    - mv "$DEST_FILE_NAME_DATE" my_nice_pdf/

artifacts:
  paths:
    - my_nice_pdf
  expire_in: 12 month
only:
  changes:
    - content/HEADER.YAML
    - content/INDEX
    - content/content.md
```