

# SIGIL: CREATE QUALITY EBOOKS ON ANY OPERATING SYSTEM

**MARCO FIORETTI**

Learn how to use the ePUB Open Standard to carry any text you want in your pocket

**WHY DO THIS?**

- Produce portable, good looking, easy to use ebooks on any operating system
- Reformat any kind of content as ebooks, to always carry it with you
- Learn by doing an open ebook standard, reusable with any other software

**E**books, that is literary works distributed not as bound stacks of paper sheets, but as digital files in the right formats, are terribly convenient.

You can back them up, carry thousands of titles in your pocket, publish them worldwide at nominal costs, and above all process and reuse their content in many ways. Ebooks are useful for everybody from teachers to corporate executives, not just bestselling authors: reformatting personal notes, company memos, courseware or generic web pages as ebooks can make all that stuff much more usable for both their authors and all their potential users.

In practice, as we hinted right at the beginning, this is true only if those ebooks are in the right formats. By this we mean Open Standards conceived specifically for ebooks, that is optimised for those paper-like screens called ereaders, but usable on any other device, of any size and form factor. "Right" also means formats that are highly structured internally, and therefore easy to write, parse and reuse with as much (Free) software as possible.

In case you hadn't noticed, this excludes the ubiquitous PDF, which is still mostly used as a picture of the printable parts of a document. The international open standard called ePUB (<http://idpf.org/epub> – see box) seems a much more sensible option for ebooks. This is why we publish this tutorial.

The multi-platform Free Software tool Sigil (<https://code.google.com/p/sigil>) is an ePub editor and formatter. Its development is currently stalled, but as long as the software installs and runs without problems, it remains one of the best ways around to not just publish ebooks, but to learn ePUB

by doing. Sigil can teach you how to produce well structured, good-looking ePUB files compatible with most ereaders around. Let's see how.

**Main concepts and user interface**

Sigil can import content in TXT, HTML or ePUB format. Whatever the input format, Sigil immediately converts and saves it as ePUB. While you may write ebooks from scratch in Sigil, you really shouldn't. Regardless of the development issue, we think it is much better to only use it to format content already written with other tools, which are probably more complete as editors, and would make it much easier to convert your work also in other formats.

To use Sigil (and ePUB in general), you only need a basic understanding of HTML and CSS markup. As a minimum, this knowledge will greatly help you when it's time to remove some code inserted by Sigil, as you'll see later.

The Sigil interface has three main tabs, which can be rearranged in several ways, or detached to independent windows: Book Browser on the left, Table of Contents (ToC) on the right, and the actual editor, which supports tabs and can work in "Book view" or "Code view" (the HTML source) in the middle.

Book Browser shows the internal structure and components of the ePUB file, whereas the ToC displays the structure of the ebook text.

Two other parts of the Sigil GUI you need to know about are the Preference Panel (Edit > Preferences) and the Clips toolbar. The most important tab of the former is the one called "Clean Source". That's where you tell Sigil when and how to clean up the imported HTML code.

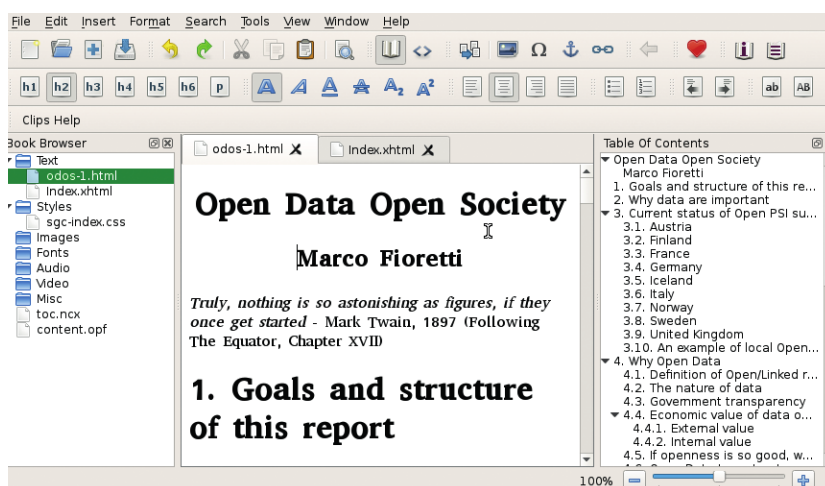
You cannot skip that step, because the HTML export filters of many programs, especially word processors like Libre Office, are unnecessarily heavy. In an attempt to produce web pages that look exactly like the original formatted text, they introduce a lot of tags that are totally useless in ebooks.

The other preferences you can set are fonts and colours of the editor, interface language, dictionaries and keyboard shortcuts.

The Clips are user-defined snippets of frequently used HTML code (one example would be CSS attributes to colour links). You can select and insert clips with a right-click in the editor window.

Many other functions of Sigil are pretty much the same as normal HTML editors, or simple word

Sigil was born as, and still is, an ePUB editor. That is why the plain editing functions get the most space in its toolbars. The real power, however, lies in the tools that generate metadata, indices and other components.



processors. We will not describe them here, because they are very intuitive we want to focus on the real value of Sigil, which is how it helps you to improve the quality and usability of ePUB files.

## First, structure your book

Metadata, that is “data about data” is what helps you and everybody else, including search engines and any other software, to make sense of your ebooks. Your ebook can only be indexed if it has the right metadata, for example.

That’s why the first place to work on a new ebook in Sigil is its Metadata Editor. You must, as a minimum, define Title, Language and Author(s). After that, the “Add Basic” button opens a menu with the 30 most common metadata types. Sigil can manage all the hundreds of metadata variables defined in the ePUB standard, and set contributors roles that go from co-author to calligrapher or censor.

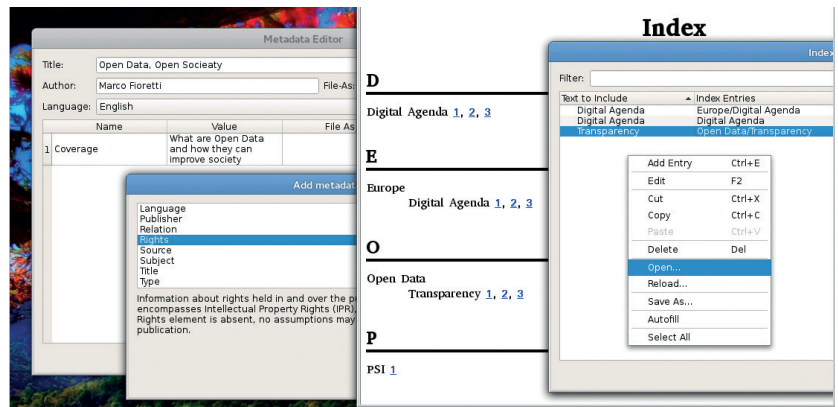
As far as the standard itself is concerned, the whole content of an ePUB book can stay in one HTML file. It is much better, however, to put all chapters into separate files, which will load faster both in Sigil and in many ereaders. You want to do it as soon as possible, to minimise the number of broken internal links you may have to fix later.

To split an ebook source into separate HTML files, put the cursor at the right point, then click on the “Split At Cursor” button. Should you change your mind, you can merge files: select them in the Book Browser, then right-click and choose Merge.

## Table of contents

The ePUB format specifies how to write standard TOCs (Table of Contents) that all ereaders can recognise and make accessible to their users, to quickly move around a book via dedicated menus or other special systems.

If the HTML source initially loaded in Sigil already has all its section headings labelled with their standard HTML markup (`<hN>...</hN>`), one click in the right place will do the job. Otherwise, select every text you want to become a section heading of a certain level in the ToC, then click on the corresponding “Hn” button. You can mark images in the same way if you want a section to start with them.



When you are done, click on the “Generate Toc” button to create the standard ePUB TOC, which will be saved in the **toc.ncx** file. You can edit this table by going to Tools > Table of Contents > Edit Table of Contents, but remember that any change done in that way will not be applied to the actual text in the source, and will be lost the next time you generate the ToC.

It’s often a good idea to placing a copy of the same ToC inside the actual content of the book, either for stylistic reasons or to make it usable even on ereaders or software programs that, for whatever reason, can’t read the **toc.ncx** file. Select Tools > Table Of Contents > Create HTML Table of Contents to get this ToC copy in a new source file, called **TOC.xhtml**, with its own CSS stylesheet (**sgc-toc.css**), then drag and drop it where you want it to be in the book.

## Indices

Besides a ToC, the other feature that makes any non-fiction book much more usable is a good index. You can define specific occurrences of strings to index, or tell Sigil to index all the occurrences of the same strings. After selecting some text, click on Tools > Index > Mark For Index to achieve the first result, or Tools > Index > Add To Index Editor for the other. You can also add entries directly to the Index Editor or (even better) load in it lists of strings to index, previously saved in plain text files.

Unless you specify different texts for them, Sigil will create entries that are exactly the strings you told it to search – those shown in the Index Editor as Text To Include. You may also use regular expressions there.

This mashup shows the Sigil Metadata Editor (left) and what you get from Sigil (centre) when you tell it to index simple or hierarchical entries (right).

## LV PRO TIP

If you plan to get serious with ebook publishing, find an HTML cheatsheet and a CSS tutorial for beginners and keep them on your desktop. As soon as you start using Sigil, you’ll need them. Besides, you can reuse the same information to design web pages!

## What does an ePUB file look like?

In order to understand what Sigil does and why, you have to know at least the general architecture and main components of the Open Standard for digital publication called ePUB (<http://idpf.org/epub>). It is not supported by all the ereaders you may find, but it is common enough that converters from ePUB to any other ebook format abound. Knowing the inside of ePUB is also essential if you plan to create or process ebooks with any other program

In extreme synthesis, an ePUB file is nothing but a compressed Zip archive of all the components of an ebook, which are given standard names and

locations. Sigil is designed for ePUB 2, but also supports some ePUB 3 features such as audio and video. If you unzipped an ePUB file, you would find in it one file and two folders. The file is simply the MIME type of the whole archive. The META-INF folder hosts a sort of pointer file, called container.xml, to the actual ebook. This is all inside the other folder, whose name is OEBPS (Open eBook Publication Structure). What the Sigil Book Browser shows, as you can see in Figure 1, is just the part of the OEBPS structure that Sigil supports.

The Table of Contents is at the top of the

hierarchy, inside the **toc.ncx** file (the extension means “Navigation Center eXtended”). The **ncx** format is obsolete in ePUB 3, but it should remain usable for a long time.

All the metadata go into the XML document named **contents.opf**, which also hosts a “Manifest”, that is a list of all the files used in the ebook.

Each category of content has its own subfolder, namely Text, Stylesheets, Images, Fonts, Audio, Video, plus Miscellanea for everything else. The text sources are normal (x)HTML files that you may open in any Web browser.

## Documentation and support

If you want to use Sigil for anything but really basic formatting, you have to begin outside it: read and keep at hand any of the HTML markup cheatsheets and "CSS for dummies" tutorials you can easily find online.

Sigil itself has a great documentation. The official User Guide is very complete and well structured: more than 35K words, which are mostly tutorials on specific issues. Besides, since it was written in Sigil, the User Guide is a great

real world example of how to use this tool: download the ePub version and load it in Sigil to see by yourself how its developers produce ebooks with it.

When the Guide isn't enough, visit the Sigil Forum at MobileRead ([www.mobilerread.com/forums/forumdisplay.php?f=203](http://www.mobilerread.com/forums/forumdisplay.php?f=203)). Among the many threads there, we recommend the one titled "Best Pre-Sigil word processor tool/workflow?", and all those that discuss Regular Expressions in Sigil.

### LV PRO TIP

These days no book, digital or in paper, is really complete and usable if it cannot be easily indexed and classified by computers. Never release an ebook if you haven't filled it with good metadata. It may be boring, but it's vital and really easy with Sigil.

You can also tell Sigil to create multiple entries for the same string and/or hierarchical ones, with the several levels separated by slashes, as in "Free Software/Linux/Ubuntu".

To actually create the index once you have finished defining its content, select Tools > Index > Create Index. The result will be saved in alphabetical order in a new page called **index.xhtml**, with its own stylesheet (**sgc-index.css**). You can edit it, but any change will be lost the next time Sigil regenerates the index.

At the source code level, indexing a word makes Sigil give it an anchor with a special class (**sigil\_index\_marker**). That's important to know, because to stop some specific occurrence of that word from appearing in the index, you must manually remove those tags from around it.

To see which snippets of text are currently indexed, switch to Code View or (much better, in our opinion), give the **sigil\_index\_marker** class a different colour in the stylesheet.

### Cross-references!

The last thing you need to make the difference between a generic, unhelpful flow of text and a really easy to use one is internal links, for notes and other cross references. To make any point in the text an anchor, that is, a destination of such links, select it,

then click on the Anchor button and give it a proper name in the pop-up window. Here, "proper" means whatever you want, as long as it begins with a letter, is unique to the whole book (so that if it would remain unique, even if you later moved that text to another file of the same book) and you use a consistent naming scheme.

To create a link to an already existing anchor (which may also be a chapter heading), click on the point where it should go and select Insert > Link. You will be able to select as destination any of the valid targets in the current ePub file, or an external URL.

This procedure is also usable for "reverse linking", that is to let readers return to whatever part of the text they were previously reading with one click, even on ereaders that lack a built-in "Back" button. You just have to invert the target and destination points. However, if you really need reverse links for many anchors, it may make more sense to add them automatically with a script.

Once you're happy with the structure of your ebook, you can start worrying about its look, comforted by the fact that writing ePub ebooks is much like writing textual content for the web. To begin with, if you write the text outside Sigil, you should carefully avoid the bad habits you may have picked up using word processors, such as adding blank lines here and there, or manually formatting text instead of using styles. This advice alone could save you lots of time and frustration when you lay out your book in Sigil.

Speaking of styles, ePub, and consequently Sigil, use the same CSS stylesheets as web pages. In case you've never met CSS before, here is a real quick copy of samples of what it can do, and how, in an ebook.

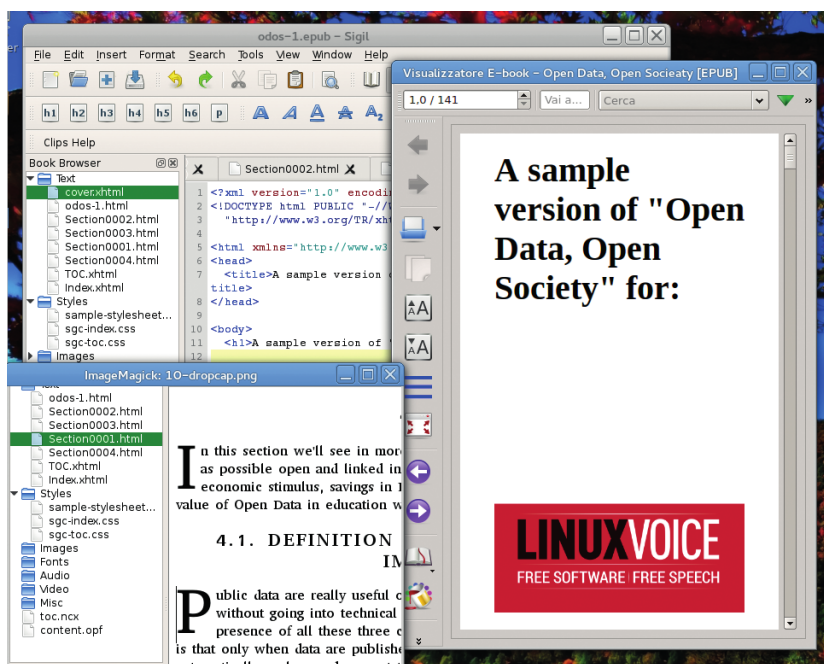
This snippet of CSS code:

```
p {
padding: 0;
margin: 0;
text-align: justified;
}
```

means "justify, with null padding and margins, all the paragraph elements (that is, all those between pairs of "<p>" and "</p>" markers in the HTML source). Drop Caps, while not supported by all ereaders, work in the same way. Adding code like this to your stylesheet:

```
span.dropcap {
float: left;
font-size: 4.7em;
line-height: 0.8em;
```

Content and structure come first, but looks are important too. Sigil supports cover design and drop caps via standard CSS stylesheets.



```
margin-right: 3pt;
```

```
margin-bottom: -0.1em;
```

```
}
```

will make a drop cap of any letter marked with that attribute in the Code View of Sigil:

```
<p><span class="dropcap">I</span> am a Drop Cap</p>
```

To add an existing CSS file to your current book, select it in File > Add Existing Files. Then, to associate it to the sources files, right-click on them in the Book Browser, select "Link Stylesheets" then tick the stylesheet you want.

CSS stylesheets are also the place to tell your ebook to use custom fonts. To be usable, the font files must have first been saved in the corresponding subfolder of the Sigil Book Browser. If you care about maximum compatibility with all ereaders the formats to use are those called OpenType or TrueType, which have the **.otf** and **.ttf** extension. Make sure to choose fonts whose licence does allow you to use them freely!

Once the fonts are in place, assign them to a CSS style, and use it in the HTML files:

```
@font-face {
```

```
font-family: 'myfont';
```

```
font-weight: normal;
```

```
font-style: normal;
```

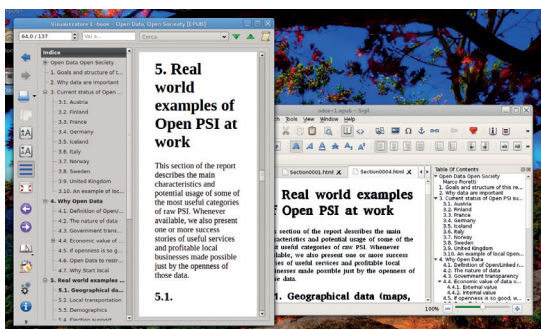
```
src: url('../Fonts/myfont.ttf');
```

```
}
```

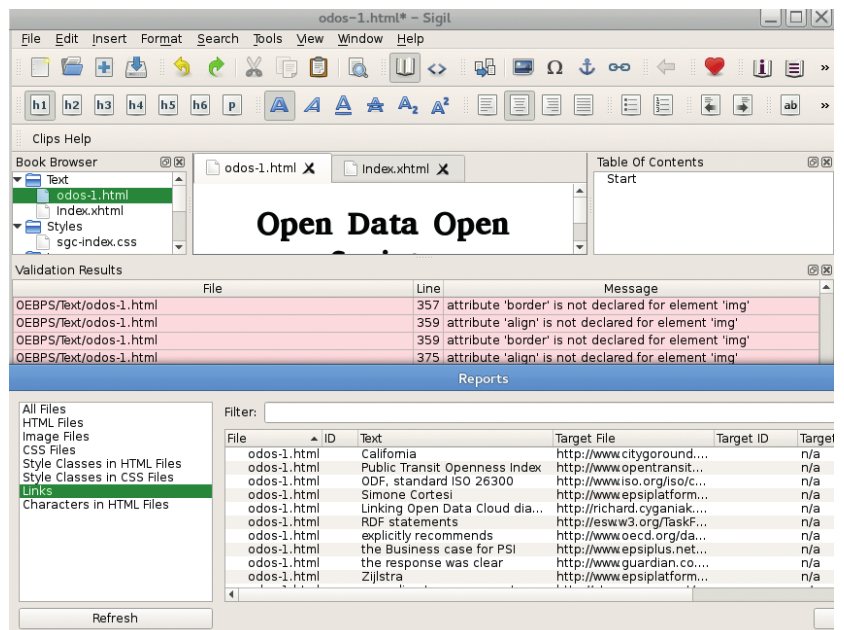
## Cover and other graphics

Adding a cover with Sigil is as easy as it gets. Select Tools > Add Cover, find the image you want to use, load it and you're done. Sigil will take create a cover source file (**cover.xhtml**) containing your image, marked up to be resizable and usable on most e-readers. Apart from images, if the (dull) default cover template that Sigil provides bores you, you have two possibilities. The quick and dirty solution is to just open the **cover.xhtml** file right there, in Sigil, and modify it as you please. If you plan to do more than one ebook with the same cover style, however, it would be better to create your own cover template.

To do this, copy the file that Sigil created to the Sigil Preferences Folder, which is **\$HOME/.local/share/sigil-ebook/sigil/**, then edit it as needed. In doing that, you can use the Sigil variables that define the path (inside the ePUB archive, not on your computer), width



Sigil can generate a standard Table of Contents that any ebook reading software understands.



and height in pixels of the cover image. They are called **SGC\_IMAGE\_FILENAME**, **SGC\_IMAGE\_WIDTH** and **SGC\_IMAGE\_HEIGHT**.

## Finally, check the result

To work as expected, any ePUB file must meet the minimum quality standards defined in the specification, and be free from internal dead links and other common errors. One of the best features of Sigil is the way that it helps you find these problems.

It doesn't hurt to run the checks that Sigil provides (Tools > Validate) as soon as you import some content. This will give you an idea of how much work, and of which kind, may be ahead. Stylesheet validation for example, and should be done as soon as possible, so you don't waste time with a layout that may look great in Sigil, but not be portable.

You can check individual stylesheets by right-clicking on them in the Book Browser and selecting the Validate option.

The F7 key starts the copy of the ePUB validator called FlightCrew (<https://code.google.com/p/flightcrew/>), which is distributed with Sigil. The reports generated directly by Sigil might be even more useful than these validators. They are great, for example, when it comes to spotting images, anchors, CSS classes and whole stylesheets that are in the ePUB file, but are never actually used. Sigil can also warn you if any of the reverse links don't actually link to each other.

Sigil is perfect for creating ePUB ebook templates. Once you have manually crafted one ebook in Sigil just like you want, reusing its files as bases for other books with the same style, via shell scripts, will be easy. But that's a challenge for another day! 📖

FlightCrew and the other validation and reporting tools included in Sigil provide a complete view of everything that may be wrong in your ebook.

### LV PRO TIP

Ebooks cluttered with wrong links or useless elements, be they images, stylesheets or unused cross-references, are bigger than necessary and harder to read and manage. Find and remove all that dead weight with the reports provided by Sigil.

Marco Fioretti is a Free Software and open data campaigner who has evangelised FOSS all over the world.