

# From Print to Ebooks: A Hybrid Publishing Toolkit for the Arts

Digital Publishing Toolkit Collective<sup>a</sup>, M. RIPHAGEN<sup>b 1</sup>,  
M. RASCH<sup>b</sup>, F. CRAMER<sup>b</sup>

<sup>a</sup>*Amsterdam University of Applied Sciences, Amsterdam, the Netherlands*

<sup>b</sup>*Institute of Network Cultures and PublishingLab, Amsterdam, the Netherlands*

**Abstract.** This article is an excerpt of the outcome of a two-year research and development project on hybrid publishing. The DPT Collective [1] developed a *Toolkit* which consists of the publication *From Print to Ebooks: A Hybrid Publishing Toolkit for the Arts* [2] and an online software kit [3] – which is meant for publishers who publish visually oriented books in mostly smaller print runs. This Toolkit focuses particularly (but not exclusively) on EPUB3 as an electronic publication format, and on Markdown [4] as a word processing format. The recommendations stem from our practical experience in collaborating on electronic publication projects with four Dutch art, design and research publishers: BISPublishers, Valiz, nai010 uitgevers and the Institute of Network Cultures .

**Keywords.** EPUB, hybrid publishing, markdown, pandoc

## 1. Introduction

‘You must change your life’ – borrowing from the philosopher Peter Sloterdijk [5], this could be the summary of our message to art-oriented and design-oriented publishers, writers, editors and designers who are currently transitioning from traditional book making to electronic publishing or – more typically – hybrid publishing of print and electronic formats. Hybrid publishing will sooner or later confront them with the need to re-think traditional publication formats, editorial and production workflows, as well as distribution opportunities.

Having said that, there are exceptions. Workflow changes can be also minor for publishers who already do all their editorial work in highly structured digital document formats such as XML (acronym for Extensible Markup Language, a version of HTML based on XML) or databases (an organized structured collection of information: common examples are address books, library catalogues and retail inventories); but this is typically only the case in scientific and technology-oriented publishing. Changes may also be minor for larger publishers who can afford outsourcing. Generating an electronic (digital) publication in parallel to a printed publication is then simply a matter of paying an external service provider, such as a document engineering company or a media design agency, to turn a Microsoft Word or Adobe InDesign file into an ebook. This process

---

<sup>1</sup>Corresponding author. E-mail: margreet@networkcultures.org.

can be quick if the book is visually simple – such as a novel or a textbook with few illustrations – and economically worthwhile if many ebooks will be sold.

We propose an alternative to the process mentioned above. Neither a complex internal IT infrastructure, nor costly outsourcing will be viable solutions for art and design oriented publishers. Unfortunately, there is no 'magic' software button that will turn a print book design into an electronic publication just like that. Since the two media are so different, each with its own specific editorial and visual design needs, such a button is unlikely to materialize in the future either. Hybrid publishing will ultimately require changes in the way the editorial work is done. The good news is that such change is possible. This research includes instructions on how to deal with the many issues that arise when making the transition from traditional to hybrid or electronic publishing.

We are not claiming that all ebooks will follow, or indeed should follow this path. We are simply laying out one of the many directions ebook creators can now take with their publication, by using simple and inexpensive tools, and without having to buy into the industry's gloss promises of multimedia and interactivity.

## **2. State of the Art**

In this section we consider a more in-depth reading of industry promises vs. reality, the basics of a text, and three levels of electronic publishing.

### *2.1. Industry Promises vs. Reality*

For art and design publishers, the challenge of 'going electronic' with their publications is greater than that faced by other fields of publishing, for a number of reasons:

- Visually oriented publications are still more difficult to realize technically in the electronic medium, particularly when designing for a multitude of different reading devices and ebook platforms.
- Small publishers are under a great deal of pressure to keep project costs low, often due to smaller budgets.
- In order to make the investment in an electronic publication durable, electronic publications must be sustainable: they should not require constant investment in technical maintenance and version updates.

There is a stark contrast between the fanciful promises of the computer industry and the often harsh reality of the new digital medium. On one hand, publishers, editors, designers and artists tend to overestimate the interactivity and multimedia possibilities of electronic publishing. These extra possibilities do exist, but in most cases bring with them higher development costs and remain specific to one particular technical platform.

On the other hand, publishers tend to underestimate how even technically simple and seemingly trivial types of electronic publications can in fact lead to a re-thinking of established publishing practices and formats. When traditional publishing formats are replaced by electronic formats, there is a real possibility for transformation. Once the book becomes electronic or hybrid, the permanence, immutability and stability typical of physical books is likely to mutate into dynamic, modular, and participative forms. Such publications can greatly benefit from the networked environment in which ebooks exist.

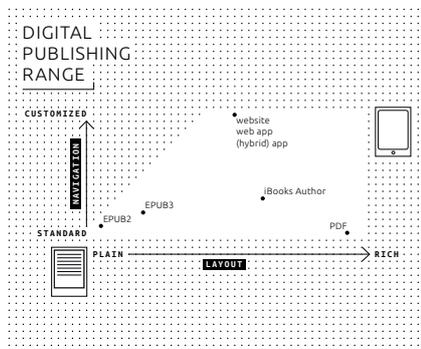


Figure 1. Digital publishing range.

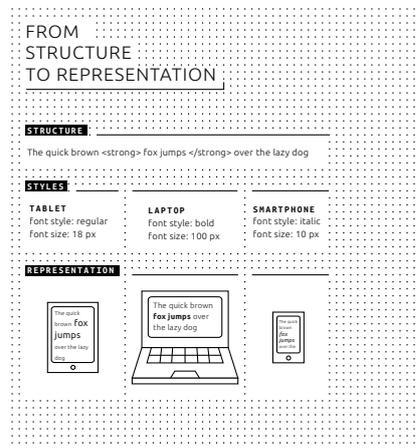


Figure 2. From structure to presentation.

Various types of electronic publications may be subject to different kinds of change. Still, the change will always be radical. For instance, an exhibition catalog can be split up into interrelated micro-monographs, which readers can download and read as individual ebooks. An ebook can be assembled from a variety of sources selected by individual readers, as is currently the case with Wikipedia, where visitors can compile their own collection of Wikipedia articles and export this compilation to an EPUB or PDF.

Users can choose from a multitude of hardware e-reader devices and software e-reader applications. The possible combinations of software and hardware are complex and virtually unlimited (see Fig. 1). This requires a certain adaptation level of not only the publisher, but also the designer. The possibilities for change can go beyond the rethinking of existing publishing formats, eventually even redefining what a book actually is.

## 2.2. The Basics of a Text

Throughout the historical development of writing, characters beyond the basic alphabet have played an increasingly important role, starting with blank spaces between words, then punctuation marks, and later markup [5] for formatting. In the electronic processing of texts, hierarchical ordering of words into sentences, sentences into paragraphs and so on, as well as additional reading aids such as bold or italic text, is made possible by using specific formatting codes. This process is called *markup* and the codes are called markup elements. All these markup elements require stable definitions and clear relationships if they are to be of any use. In order to establish which markup is allowed and how it should be used, markup languages were defined.

This is especially important and relevant in the context of hybrid publishing because it makes it possible, at the later stage of visual design, not only to define how each of these markup elements will be displayed, but also to provide different definitions for each specific output. For example, we can decide that for output A (say the printed book), text marked as 'chapter heading' will be centered on the page, in a different font and larger font size than the running text, while text marked as 'quotation' will be rendered in the same font and size as the running text, but in italics; while for output B we could instead decide to render chapter headings as bold and quotations as underlined text. By

combining the structured text with a different style sheet for each output format, a variety of end products can thus be generated using one single structured text. However, in order to make this possible, the source text must be as systematically structured as possible (see Fig. 2).

In electronic texts, markup has developed into two general types: 'What You See Is What You Get' [7] or visual-typographic (as in the markup tag 'bold type') vs. logical-semantic (as in the markup tag 'emphasis'). The logical-semantic markup is the foundation of hybrid publishing, since it can be translated into whatever visual formatting is most suitable for each particular medium. One of the main advantages of electronic books is that the same content can be published in a wide variety of formats.

All these new possibilities will require thorough, and potentially more labor-intensive, editorial and production strategies. Not only because of the possibility of representing the same content in a variety of forms, but more importantly because once they are properly edited and stored electronically, the content and its constituent parts can be endlessly used and re-used in different ways, now and in the future. This also means that electronic publishing will, in most cases, not bring any significant decrease in production costs. Though if one produces a series, you can benefit from for example, earlier made paragraph styles and predefined structures.

### *2.3. Genres of Publication*

In the present research project we deal with a variety of publication products. In art and design publishing, the most common genres are: research publications, art/design catalogues, artists'/designers' books, and art/design periodicals [12]. The opportunities and challenges of electronic publishing are different for each genre. Common opportunities include searchability, ease of access and distribution, and modularization of content; common difficulties include layout consistency, page numbering and referencing, and potentially large file sizes. Independent of the different genres listed above, we can distinguish three levels of electronic publishing:

1. One-to-one, where the book is one single product published in different media.
2. One-to-many, where the book has different appearances in different media.
3. One-to-database, where the book is based on the content of a database which can be used in a number of ways.

The scope of electronic publishing ranges from the simple conversion of a paper book, to an electronic publication (for example, a PDF of the print edition as an ebook), to full-scale electronic publications which incorporate advanced digital formats such as video, or are published as 'native apps' (applications developed for a particular platform or device).

## **3. Towards a Hybrid Workflow Based on Markdown**

Creating a workflow that is both structured and flexible enough to cater a variety of demands is a key step towards establishing an efficient electronic or hybrid publishing strategy. What we propose here is a hybrid workflow based on the need for publishing across different media, while keeping the main part of the work process in-house rather than outsourcing it.

### *3.1. Electronic Publishing Workflows: Desktop Publishing and Markdown*

Instead of developing a digital publication based on the printed book at the end of a production process, as is common practice by publishers, the main workflow should be adapted at an earlier stage, and made efficient and practical for hybrid publishing. So rather than working separately on the PDF for the print book, the EPUB version, and a Kindle edition, the workflow is instead focused on a single source file (in the Markdown format) which can easily be converted into these different output formats using a relatively small number of digital tools.

#### *3.1.1. Desktop Publishing (Traditional) Workflow (From Word to InDesign to Ebook)*

A brief description of the desktop publishing (DTP) workflow currently used by many publishers would be: a Microsoft Word file is imported into InDesign and, after designing and editing, exported to PDF, ready to be printed. After work on the printed edition has been completed, the book may be converted into an electronic version which follows the design of the 'original' as closely as possible. This traditional, print-oriented workflow can be seen as the standard for one-to-one publications (see Fig. 3). Advantages of this workflow are that it is simple and linear, and there are no version branches. You end up with one consolidated manuscript, and WYSIWYG when it comes to design. The main disadvantage of the DTP workflow in 'going electronic' is that it is focused on one single medium, and that the steps to go from there to a digital edition are quite laborious and do not make full use of the possibilities offered by electronic publishing.

#### *3.1.2. From Microsoft Word (.docx) to EPUB*

Desktop publishing applications such as InDesign, and WYSIWYG word processors such as Microsoft Word or OpenOffice, are generally not well suited for processing structured text (see section 2.2). Though it is possible to work in a structured manner, for example by using style definitions rather than manually applying formatting, the user is not required to make a distinction between formatting and structure, which is essential in the world of digital publishing. In order to obtain the best possible EPUB file, the .docx file should be formatted using only Word's standard paragraph styles such as 'Normal', 'Title', 'Subtitle', 'Quote' and most importantly 'Heading 1', 'Heading 2', 'Heading 3' for the headings according to their logical hierarchy. For example: 'Heading 1' for chapters, 'Heading 2' for sections, 'Heading 3' for sub-sections. Since the resulting EPUB document will contain a table of contents and document navigation menu based on the 'Heading' hierarchy, a proper structuring of headings is crucial. Word footnotes will appear as linked endnotes in the EPUB, thereby elegantly simplifying an otherwise tedious document redesign task.

Word unfortunately lacks two features that would make it more suitable for hybrid publishing projects:

1. Word does not have a 'strict mode' that would 'force' all writers and editors of a document to use only defined paragraph styles instead of manual formatting.
2. Word provides no automatic or semi-automatic tools for finding manual formatting and replacing it with predefined paragraph styles. The only way to achieve this is to manually review and adjust the whole document.

Often, such inconsistencies in a Word document will only become visible after the EPUB conversion. We recommend two ways of working with Word + Pandoc (an open-source document converter).

1. Conversion from Word to EPUB using Pandoc directly from the Terminal (Mac) / Command Prompt (Windows), or using the browser-based converter developed for this Toolkit [8]. This should only be done only at the very end of the editorial process, because after this conversion no further editorial changes can be applied to the Word document.
2. Conversion from Word to Markdown using Pandoc. Since Pandoc can also convert files *to* the Markdown format, this is often the preferable option, especially for complex publishing projects. The resulting Markdown file can then be used as the 'master file' for conversions to various other file formats (such as EPUB, or HTML for publishing on a website). The advantage of converting to Markdown is that any formatting glitch in the Word document will now become clearly visible. For example, a heading formatted manually will be converted as **\*\*heading\*\*** while a heading formatted with the proper style definition will be converted as **#heading**. This makes it much easier to clean up the internal formatting of the document and produce a 'clean' master file for all subsequent document conversions. The EPUB generated from this Markdown file will in most cases be much better structured than an EPUB directly generated from the Word file, making the subsequent work to be performed by the designer much easier.

We would advise against using Pandoc to convert back and forth between Word and EPUB. If the Word document is subject to further editorial changes, then the conversion to EPUB (as in the first scenario) should be done again, as would any work already done by the designer on the previously exported EPUB file. Therefore, if possible the editorial changes should be implemented directly in the EPUB or Markdown file.

### 3.1.3. *Cleaning Up Markdown*

Since Markdown is not a word processing application but a document format, it does not provide functions such as automatic renumbering of footnotes and list items. In fact, such numbers don't matter since everything will be renumbered during the document conversion anyway.

However, Pandoc can be used to 'clean up' the Markdown source text; the trick is to convert the document *from* Markdown *to* Markdown. Open a Terminal window (Mac) or Command Prompt window (Windows) and type the following line (be sure to first place the file in the appropriate folder and to navigate to that folder).

```
pandoc beowulf.md -f Markdown -t Markdown -o beowulf_clean.md
```

This means that the program Pandoc is instructed to convert the file 'beowulf.md' file from Markdown ('-f Markdown') to Markdown ('-t Markdown') and save the (cleaned up) result in a new file called 'beowulf\_clean.md' ('-o beowulf\_clean.md').

### 3.2. *Markdown Workflow*

As mentioned above, we recommend the use of the markup language Markdown as part of a hybrid workflow. Though Markdown is not perfect, it is much easier to work with

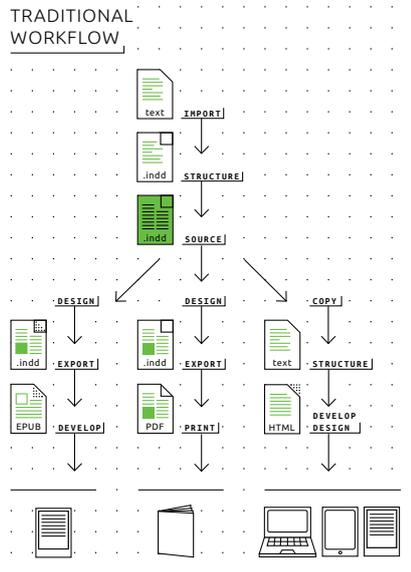


Figure 3. Traditional workflow.

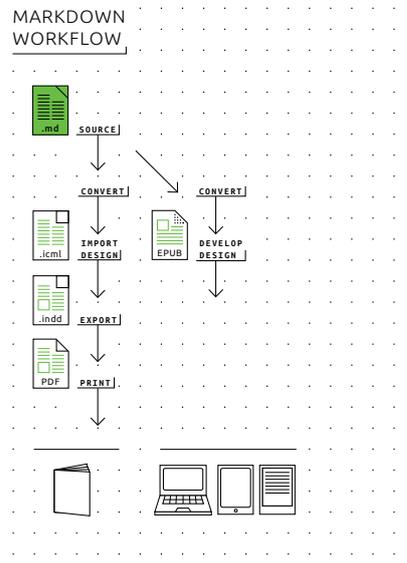


Figure 4. Markdown workflow.



Figure 5. Alice's Adventures in Wonderland.

than, say, the complex markup language XML. Markdown allows for the creation of structured texts, an important requirement in hybrid publishing (see Fig. 4).

### 3.3. Introduction: Advantages and Limitations

John Gruber, the creator and main developer of Markdown, describes Markdown on his website as follows: 'Markdown allows you to write using an easy-to-read, easy-to-write plain-text format, then convert it to structurally valid XHTML (or HTML).' [6] In other words, Markdown is a way of formatting plain text using human-readable formatting symbols, rather than HTML-style tags such as '<b>' for bold or '<h1>' to define a top-level heading. For example, this is what the beginning of *Alice's Adventures in Wonderland* would look like in Markdown (see Fig. 5).

In this example, the tag '# ' defines a top-level heading, '## ' a second-level heading, '\_ ' italic text, '\*\* ' bold text, and '>' a block quote. Markdown also provides tags for defining lists, embedded images, and links. The popular extended version Multi-Markdown (an extension of the markup language Markdown, with additional support for

footnotes, tables, etc.) provides further support for footnotes, tables, mathematical formulas, cross-references, bibliographies and definition lists. Using simple Open Source conversion applications such as Pandoc, text formatted using Markdown can be automatically converted to well-structured HTML, EPUB, PDF, RTF or other document formats, requiring no manual adjustments.

Markdown is a product of Internet culture. It uses ad-hoc formatting signs commonly used in e-mail and chat platforms, and further popularized on blogging platforms, to provide a standardized, human-readable, user-friendly and well-structured document format, suitable for long-term storage and as a basic source for conversions to contemporary and future document formats. While its formatting *syntax* is simple, it is also both strict and unambiguous enough to allow multiple writers and editors to work on a single document without unnecessary confusion. Another advantage of Markdown is that it can be written and edited in any software application capable of processing basic text: unlike the proprietary file formats of Microsoft Word or other word processors, Markdown can be opened using a simple text editor.

Why do we recommend Markdown in particular? For certain publishing projects – for example, handbooks or books derived from wikis, it may be worth considering alternatives to Markdown, such as reStructuredText. There are many text editing and document conversion programs that support Markdown, such as Mou or MacDown.

However, Markdown/MultiMarkdown is not a magical one-size-fits-all solution. It is particularly well suited for text-oriented documents, but quite limited for creating visually oriented documents, and not really of much use for interactive publishing formats.

Markdown and similar formatting/markup languages are designed for workflows in which there is a clear separation between editorial work (involving writers, translators and editors) and publication design. For publications requiring extensive interaction between writers/editors and visual designers/artists from the very beginning of the authoring process, other tools and workflows are preferable.

### 3.3.1. *Markdown vs. XML*

XML is designed for creating structured documents with a clear separation between logical structure and visual formatting. It is the most detailed structuring and formatting language ever developed, and provides the foundation for many other such languages. For example, both HTML and Microsoft Word's .docx are XML-based document formats. So why shouldn't we use XML then? While XML theoretically provides an ideal way of working with single-format files to produce multiple output formats, we do not recommend it for small, independent publishing houses.

The main reason for this is that the broad versatility of XML adds several layers of complexity. Markdown on the other hand can easily be used by non-technical users while still providing good structure and better document conversion into HTML, EPUB and many other formats than Microsoft Word and similar word processing applications. Technically speaking, Markdown provides some of the same features and advantages as XML does, namely separation of content structure from visual layout and painless conversion into multiple output formats. However, unlike XML, it cannot be extended with custom, self-defined markup tags. Still, and particularly for those just getting started with digital or hybrid publishing, Markdown should be more than adequate in almost all circumstances.

## 4. Showcases, How To Make a Simple EPUB

All showcases underneath are a collaboration with one of the above mentioned publishers, one or more designers and developers from the DPT Collective.

### 4.1. Showcase BISPublishers

Where other groups of the DPT research group have focused on developing an EPUB, the BISPublishers publication *Sketching, drawing techniques for product designers* [9] [10] has explored the possibilities of a hybrid application as medium. With Phonegap, a mobile development framework, a bridge is made between HTML and a native app, which allows for a multichannel publishing workflow.

### 4.2. Showcase Valiz

Valiz has developed an EPUB version of *Common Skin*, a publication in the series *Context Without Walls* [11]. The multilanguage publications in this series focus on artists from all over the world and contain essays as well as pictures. The transformation from a printed version to an EPUB is not a one-on-one translation, the limitations of graphic design in the EPUB format and the complexities of the production process for a series of publications had to be considered. A publication platform called *EPUBster* has been created to assist with the production of the EPUB [12].

### 4.3. Showcase nai010 uitgevers

nai010 has developed a digital version of the *Stedelijk Museum Highlights Catalogue* [13]. Instead of creating an ebook version of the existing catalogue, the group has opted to develop a digital distribution platform accessed via a mobile (web)application. The application allows users to find, filter, search, preview and collect highlights from the *Stedelijk Museum* collection in preparation for their visit to the museum. Related content (essays, interviews, images, etc.) for the selected highlights may be chosen and are subsequently bundled as a personalised EPUB catalogue, available for generation within the *My Highlights* application [14]. The generated publication may then be downloaded and viewed (offline) on any device capable of rendering EPUBs.

### 4.4. Showcase Institute of Network Cultures (INC)

The hybrid workflow INC proposes is based on the need for publishing across different media, while keeping the majority of the work process in-house instead of outsourcing. The research of the INC was aimed at making the transition from a print-centered publication process towards a digital-and-print (hybrid) publication process. The leading question being: how to handle documents in order to publish on different platforms? Creating a workflow that is both structured and flexible enough to cater for different choices is a key step towards an efficient electronic or hybrid publishing strategy. From-scratch development of each publication format is replaced by single source-multi format publishing.

## 5. Future of Publishing Culture

How will people read in the future? Do people read ebooks from the beginning to the end, or do they casually browse them - or merely keep them as searchable items in a personal electronic library? Do they read PDF versions of books and articles on display screens, or do they first print them on paper? Or will long, complex texts and printed books become fashionable again in a more mature technological environment? It is very possible that in the course of time, users will adjust to the new reading technologies, and have no trouble digesting long and complex texts from the displays of their mobile devices. Already today, increasing numbers of writers and artists self-publish their works. The role of publishers as book producers is rapidly becoming a thing of the past. However, at a certain point of overproduction and oversaturation, publishers may redefine their role to become aggregators and curators.

The electronic publishing models which we have focused on in this Toolkit, using technologies such as EPUB, may in some cases appear counter-intuitive to today's digital media culture: why create what are essentially offline websites in ZIP files, in this age of 'cloud computing' and an 'always-on' culture in which we are constantly connected and networked? However, it is precisely the ephemeral nature of networked media that makes a format like EPUB increasingly attractive. As an offline, stable medium based on World Wide Web technology, it is perfect for everyone who wishes to personally curate, collect and preserve what otherwise may soon be lost.

Faced with on one hand all these new possibilities for self-publishing and self-curating, and on the other hand the rise and consolidation of huge commercial monopolies, the art and craftsmanship of publishing will have no choice but to reinvent and rebuild itself. The changes required may well be greater and more extensive than initially expected.

## References

- [1] Digital Publishing Toolkit Collective, consists of: Marc de Bruijn, Liz Castro, Florian Cramer, Joost Kircz, Silvio Lorusso, Michael Murtaugh, Pia Pol, Miriam Rasch, Margreet Riphagen, Loes Sikkes and Kimberley Spreeuwenberg.
- [2] J. Monk, F. Cramer, M. Rasch, M. Riphagen. & Digital Publishing Collective (2014). *From Print to Ebooks: a Hybrid Publishing Toolkit for the Arts*. Amsterdam, Institute of Network Cultures, 2015.
- [3] GitHub [Internet]. [cited 2015 June 2]. Available from: <https://github.com/DigitalPublishingToolkit/>.
- [4] Merriam-Webster [Internet]. [cited 2015 June 2]. Available from: <http://www.merriam-webster.com/dictionary/markup%20language>.
- [5] P. Sloterdijk. & W. Hoban. *You must change your life: On anthropotechnics*. Cambridge, UK: Polity, 2013
- [6] John Gruber, *Markdown: Introduction* [Internet]. [cited 2015 June 2]. Available from: [daringfireball.net/projects/Markdown/](http://daringfireball.net/projects/Markdown/).
- [7] Merriam-Webster [Internet]. [cited 2015 June 2]. Available from: <http://www.merriam-webster.com/dictionary/wysiwyg>.
- [8] An overview can be found on the Digital Publishing Toolkit Software Showcase [Internet]. [cited 2015 June 2]. Available from: <http://pandoc.networkcultures.org/> or go directly to <http://pandoc.networkcultures.org/hybrid.html>.
- [9] K. Eissen, and R. Steur, *Sketching Drawing Techniques for Product Designers*. Holanda: Bis, 2007.
- [10] GitHub Sketching [Internet], [cited 2015 June 2]. Available from: <https://github.com/DigitalPublishingToolkit/Sketching>
- [11] D. Pappers, L. Levy, M. Mihindou, and P. Mason. *Common Skin*. Valiz, 2014.

- [12] EPUBster is a web application to create and edit EPUBs, written in CakePHP. Content formatted using Markdown is used as input to generate publications using the EPUB 3 file format [Internet]. [cited 2015 June 2] Available from: <https://github.com/DigitalPublishingToolkit/epubster>.
- [13] H. De Man. *Stedelijk Collection Highlights: 150 Artists from the Collection of the Stedelijk Museum Amsterdam*. Amsterdam: Stedelijk Museum, 2012.
- [14] My Highlights application [Internet]. [cited 2015 June 2]. Available from: <https://github.com/DigitalPublishingToolkit/My-Highlights>.