

A peer-reviewed version of this preprint was published in PeerJ on 8 May 2017.

[View the peer-reviewed version](https://peerj.com/articles/cs-112) (peerj.com/articles/cs-112), which is the preferred citable publication unless you specifically need to cite this preprint.

Krewinkel A, Winkler R. 2017. Formatting Open Science: agilely creating multiple document formats for academic manuscripts with Pandoc Scholar. PeerJ Computer Science 3:e112
<https://doi.org/10.7717/peerj-cs.112>

1 Formatting Open Science: agilely creating 2 multiple document formats for academic 3 manuscripts with Pandoc Scholar

4 **Albert Krewinkel¹ and Robert Winkler^{2,✉}**

5 **¹Pandoc Development Team**

6 **²CINVESTAV Unidad Irapuato, Department of Biochemistry and Biotechnology**

7 Corresponding author:

8 Prof. Dr. Robert Winkler[✉]

9 Email address: robert.winkler@cinvestav.mx

10 **ABSTRACT**

The timely publication of scientific results is essential for dynamic advances in science. The ubiquitous availability of computers which are connected to a global network made the rapid and low-cost distribution of information through electronic channels possible. New concepts, such as Open Access publishing and preprint servers are currently changing the traditional print media business towards a community-driven peer production. However, the cost of scientific literature generation, which is either charged to readers, authors or sponsors, is still high. The main active participants in the authoring and evaluation of scientific manuscripts are volunteers, and the cost for online publishing infrastructure is close to negligible. A major time and cost factor is the formatting of manuscripts in the production stage. In this article we demonstrate the feasibility of writing scientific manuscripts in plain markdown (MD) text files, which can be easily converted into common publication formats, such as PDF, HTML or EPUB, using pandoc. The simple syntax of markdown assures the long-term readability of raw files and the development of software and workflows. We show the implementation of typical elements of scientific manuscripts – formulas, tables, code blocks and citations – and present tools for editing, collaborative writing and version control. We give an example on how to prepare a manuscript with distinct output formats, a DOCX file for submission to a journal, and a LATEX/PDF version for deposition as a PeerJ preprint. Further, we implemented new features for supporting 'semantic web' applications, such as the 'journal article tag suite' - JATS, and the 'citation typing ontology' - CiTO standard. Reducing the work spent on manuscript formatting translates directly to time and cost savings for writers, publishers, readers and sponsors. Therefore, the adoption of the MD format contributes to the agile production of open science literature. Pandoc Scholar is freely available from <https://github.com/pandoc-scholar>.

11 **Keywords:** open science, document formats, markdown, latex, publishing, typesetting

12 INTRODUCTION

13 Agile development of science depends on the continuous exchange of information between researchers
14 (Woelfle, Olliaro & Todd, 2011). In the past, physical copies of scientific works had to be produced and
15 distributed. Therefore, publishers needed to invest considerable resources for typesetting and printing.
16 Since the journals were mainly financed by their subscribers, their editors not only had to decide on the
17 scientific quality of a submitted manuscript, but also on the potential interest to their readers. The avail-
18 ability of globally connected computers enabled the rapid exchange of information at low cost. Yochai
19 Benkler (2006) predicts important changes in the information production economy, which are based on
20 three observations:

- 21 1. A nonmarket motivation in areas such as education, arts, science, politics and theology.
- 22 2. The actual rise of nonmarket production, made possible through networked individuals and coord-
23 inate effects.
- 24 3. The emergence of large-scale peer production, e.g. of software and encyclopedias.

25 Immaterial goods such as knowledge and culture are not lost when consumed or shared – they are ‘non-
26 rival’ –, and they enable a networked information economy, which is not commercially driven (Benkler,
27 2006).

28 Preprints and e-prints

29 In some areas of science a preprint culture, i.e. a paper-based exchange system of research ideas and
30 results, already existed when Paul Ginsparg in 1991 initiated a server for the distribution of electronic
31 preprints – ‘e-prints’ – about high-energy particle theory at the Los Alamos National Laboratory (LANL),
32 USA (Ginsparg, 1994). Later, the LANL server moved with Ginsparg to Cornell University, USA, and
33 was renamed as arXiv (Butler, 2001). Currently, arXiv (<https://arxiv.org/>) publishes e-prints related
34 to physics, mathematics, computer science, quantitative biology, quantitative finance and statistics.
35 Just a few years after the start of the first preprint servers, their important contribution to scientific com-
36 munication was evident (Ginsparg, 1994; Youngen, 1998; Brown, 2001). In 2014, arXiv reached the
37 impressive number of 1 million e-prints (Van Noorden, 2014).

38 In more conservative areas, such as chemistry and biology, accepting the publishing prior peer-review
39 took more time (Brown, 2003). A preprint server for life sciences (<http://biorxiv.org/>) was
40 launched by the Cold Spring Habor Laboratory, USA, in 2013 (Callaway, 2013). *PeerJ preprints*
41 (<https://peerj.com/preprints/>), started in the same year, accepts manuscripts from biological
42 sciences, medical sciences, health sciences and computer sciences.

43 The terms ‘preprints’ and ‘e-prints’ are used synonymously, since the physical distribution of preprints
44 has become obsolete. A major drawback of preprint publishing are the sometimes restrictive policies of
45 scientific publishers. The SHERPA/RoMEO project informs about copyright policies and self-archiving
46 options of individual publishers (<http://www.sherpa.ac.uk/romeo/>).

47 Open Access

48 The term ‘Open Access’ (OA) was introduced 2002 by the Budapest Open Access Initiative and was
49 defined as:

50 “Barrier-free access to online works and other resources. OA literature is digital, online, free of charge
51 (gratis OA), and free of needless copyright and licensing restrictions (libre OA).” (Suber, 2012)

52 Frustrated by the difficulty to access even digitized scientific literature, three scientists founded the *Public*
53 *Library of Science (PLoS)*. In 2003, *PLoS Biology* was published as the first fully Open Access journal
54 for biology (Brown, Eisen & Varmus, 2003; Eisen, 2003).

55 Thanks to the great success of OA publishing, many conventional print publishers now offer a so-called
56 ‘Open Access option’, i.e. to make accepted articles free to read for an additional payment by the authors.
57 The copyright in these hybrid models might remain with the publisher, whilst fully OA usually provide

58 a liberal license, such as the Creative Commons Attribution 4.0 International (CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/>).

60 OA literature is only one component of a more general *open* philosophy, which also includes the access
61 to scholarships, software, and data (Willinsky, 2005). Interestingly, there are several different ‘schools
62 of thought’ on how to understand and define *Open Science*, as well the position that any science is open
63 by definition, because of its objective to make generated knowledge public (Fecher & Friesike, 2014).

64 Cost of journal article production

65 In a recent study, the article processing charges (APCs) for research intensive universities in the USA
66 and Canada were estimated to be about 1,800 USD for fully OA journals and 3,000 USD for hybrid
67 OA journals (Solomon & Björk, 2016). PeerJ (<https://peerj.com/>), an OA journal for biological
68 and computer sciences launched in 2013, drastically reduced the publishing cost, offering its members a
69 life-time publishing plan for a small registration fee (Van Noorden, 2012); alternatively the authors can
70 choose to pay an APC of 1,095 USD, which may be cheaper, if multiple co-authors participate.

71 Examples such as the *Journal of Statistical Software* (*JSS*, <https://www.jstatsoft.org/>) and *eLife*
72 (<https://elifesciences.org/>) demonstrate the possibility of completely community-supported OA
73 publications. **Fig. 1** compares the APCs of different OA publishing business models.

74 *JSS* and *eLife* are peer-reviewed and indexed by Thomson Reuters. Both journals are located in the
75 Q1 quality quartile in all their registered subject categories of the Scimago Journal & Country Rank
76 (<http://www.scimagojr.com/>), demonstrating that high-quality publications can be produced without
77 charging the scientific authors or readers.

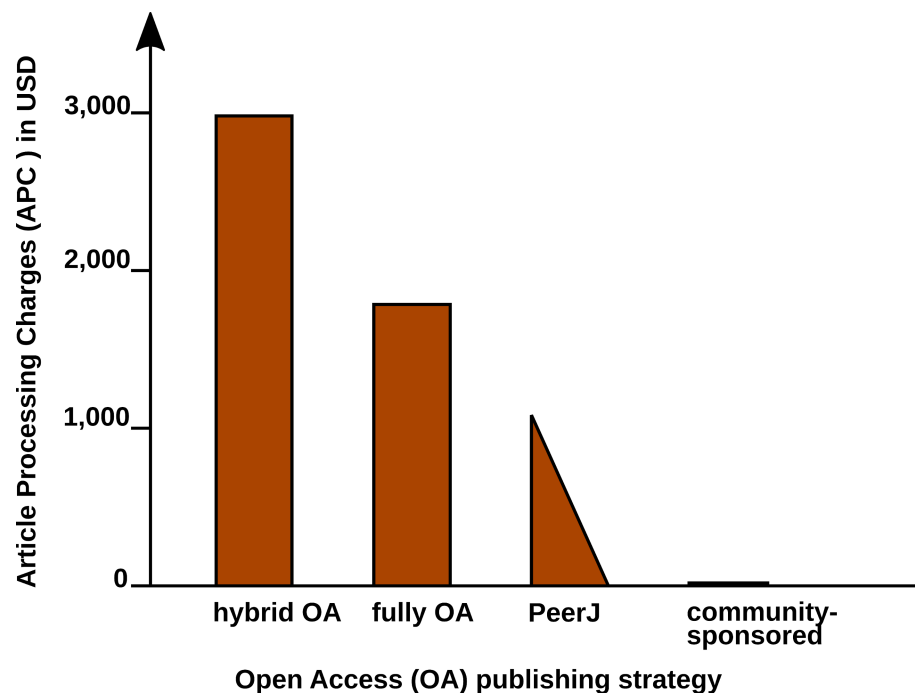


Figure 1. Article Processing Charge (APCs) that authors have to pay for with different Open Access (OA) publishing models. Data from (Solomon & Björk, 2016) and journal web-pages.

78 In 2009, a study was carried out concerning the “*Economic Implications of Alternative Scholarly Publishing Models*”, which demonstrates an overall societal benefit by using OA publishing model (Houghton
79 et al., 2009). In the same report, the real publication costs are evaluated. The relative costs of an article
80 for the publisher are represented in **Fig. 2**.

82 Conventional publishers justify their high subscription or APC prices with the added value, e.g. journal-
83 ism (stated in the graphics as ‘non-article processing’). But also stakeholder profits, which could be as

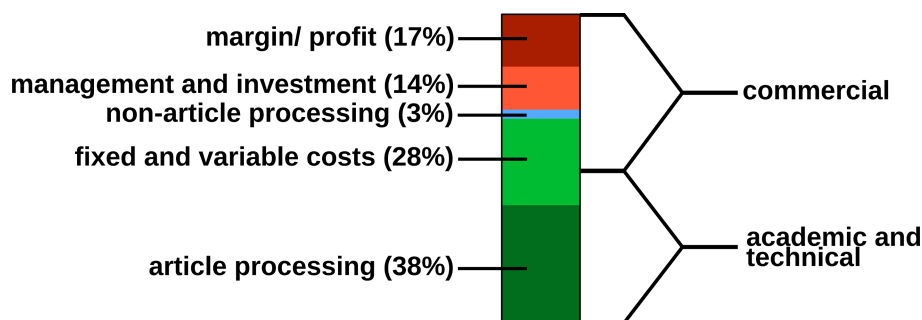


Figure 2. Estimated publishing cost for a ‘hybrid’ journal (conventional with Open Access option). Data from (Houghton et al., 2009).

84 high as 50%, must be considered, and are withdrawn from the science budget (Van Noorden, 2013).

85 Generally, the production costs of an article could be roughly divided into commercial and academic/
 86 technical costs (**Fig. 2**). For nonmarket production, the commercial costs such as margins/ profits, man-
 87 agement etc. can be drastically reduced. Hardware and services for hosting an editorial system, such as
 88 Open Journal Systems of the Public Knowledge Project (<https://pkp.sfu.ca/ojs/>) can be provided
 89 by public institutions. Employed scholars can perform editor and reviewer activities without additional
 90 cost for the journals. Nevertheless, ‘article processing’, which includes the manuscript handling during
 91 peer review and production represents the most expensive part.

92 Therefore, we investigated a strategy for the efficient formatting of scientific manuscripts.

93 **Current standard publishing formats**

94 Generally speaking, a scientific manuscript is composed of contents and formatting. While the content,
 95 i.e. text, figures, tables, citations etc., may remain the same between different publishing forms and jour-
 96 nal styles, the formatting can be very different. Most publishers require the formatting of submitted
 97 manuscripts in a certain format. Ignoring this **Guide for Authors**, e.g. by submitting a manuscript with
 98 a different reference style, gives a negative impression with a journal’s editorial staff. Too carelessly
 99 prepared manuscripts can even provoke a straight ‘desk-reject’ (Volmer & Stokes, 2016).

100 Currently DOC(X), LATEX and/ or PDF file formats are the most frequently used formats for journal
 101 submission platforms. But even if the content of a submitted manuscript might be accepted during the
 102 peer review ‘as is’, the format still needs to be adjusted to the particular publication style in the production
 103 stage. For the electronic distribution and archiving of scientific works, which is gaining more and more
 104 importance, additional formats (EPUB, (X)HTML, JATS) need to be generated. **Tab. 1** lists the file
 105 formats which are currently the most relevant ones for scientific publishing.

106 Although the content elements of documents, such as title, author, abstract, text, figures, tables, etc.,
 107 remain the same, the syntax of the file formats is rather different. **Tab. 2** demonstrates some simple
 108 examples of differences in different markup languages.

109 Documents with the commonly used Office Open XML (DOCX Microsoft Word files) and OpenDocu-
 110 ment (ODT LibreOffice) file formats can be opened in a standard text editor after unzipping. However,
 111 content and formatting information is distributed into various folders and files. Practically speaking, those
 112 file formats require the use of special word processing software.

113 From a writer’s perspective, the use of *What You See Is What You Get (WYSIWYG)* programs such as
 114 Microsoft Word, WPS Office or LibreOffice might be convenient, because the formatting of the document
 115 is directly visible. But the complicated syntax specifications often result in problems when using different
 116 software versions and for collaborative writing. Simple conversions between file formats can be difficult
 117 or impossible. In a worst-case scenario, ‘old’ files cannot be opened any more for lack of compatible
 118 software.

119 In some parts of the scientific community therefore LATEX, a typesetting program in plain text format,
 120 is very popular. With LATEX, documents with highest typographic quality can be produced. However,
 121 the source files are cluttered with LATEX commands and the source text can be complicated to read.
 122 Causes of compilation errors in LATEX are sometimes difficult to find. Therefore, LATEX is not very
 123 user friendly, especially for casual writers or beginners.

Table 1. Current standard formats for scientific publishing.

Type	Description	Use	Syntax	Reference
DOCX	Office Open XML	WYSIWYG editing	XML, ZIP	(Ngo, 2006)
ODT	OpenDocument	WYSIWYG editing	XML, ZIP	(Brauer et al., 2005)
PDF	portable document	print replacement	PDF	(International Organization for Standardization, 2013)
EPUB	electronic publishing	e-books	HTML5, ZIP	(Eikebrokk, Dahl & Kessel, 2014)
JATS	journal article tag suite	journal publishing	XML	(National Information Standards Organization, 2012)
LATEX	typesetting system	high-quality print	TEX	(Lamport, 1994)
HTML	hypertext markup	websites	(X)HTML	(Raggett et al., 1999; Hickson et al., 2014)
MD	Markdown	lightweight markup	plain text MD	(Ovadia, 2014; Leonard, 2016)

Table 2. Examples for formatting elements and their implementations in different markup languages.

Element	Markdown	LATEX	HTML
structure			
section	# Intro	\section{Intro}	<h1><Intro></h1>
subsection	## History	\subsection{History}	<h2><History></h2>
text style			
bold	**text**	\textbf{text}	text
italics	*text*	\textit{text}	<i>text</i>
links			
HTTP link	<https://arxiv.org>	\usepackage{url} \url{https://arxiv.org}	

124 In academic publishing, it is additionally desirable to create different output formats from the same source
 125 text:

- 126 • For the publishing of a book, with a print version in PDF and an electronic version in EPUB.
- 127 • For the distribution of a seminar script, with an online version in HTML and a print version in PDF.
- 128 • For submitting a journal manuscript for peer-review in DOCX, as well as a preprint version with another journal style in PDF.
- 129 • For archiving and exchanging article data using the Journal Article Tag Suite (JATS) (National Information Standards Organization, 2012), a standardized format developed by the NLM.

133 Some of the tasks can be performed e.g. with LATEX, but an integrated solution remains a challenge.
 134 Several programs for the conversion between documents formats exist, such as the e-book library program calibre <http://calibre-ebook.com/>. But the results of such conversions are often not satisfactory
 135

136 and require substantial manual corrections.

137 Therefore, we were looking for a solution that enables the creation of scientific manuscripts in a simple
 138 format, with the subsequent generation of multiple output formats. The need for hybrid publishing has
 139 been recognized outside of science (Kielhorn, 2011; DPT Collective, 2015), but the requirements specific
 140 to scientific publishing have not been addressed so far. Therefore, we investigated the possibility to
 141 generate multiple publication formats from a simple manuscript source file.

142 CONCEPTS OF MARKDOWN AND PANDOC

143 Markdown was originally developed by John Gruber in collaboration with Aaron Swartz, with the goal
 144 to simplify the writing of HTML documents <http://daringfireball.net/projects/markdown/>.
 145 Instead of coding a file in HTML syntax, the content of a document is written in plain text and annotated
 146 with simple tags which define the formatting. Subsequently, the Markdown (MD) files are parsed to
 147 generate the final HTML document. With this concept, the source file remains easily readable and the
 148 author can focus on the contents rather than formatting. Despite its original focus on the web, the MD
 149 format has been proven to be well suited for academic writing (Ovadia, 2014). In particular, pandoc-
 150 flavored MD (<http://pandoc.org/>) adds several extensions which facilitate the authoring of academic
 151 documents and their conversion into multiple output formats. **Tab. 2** demonstrates the simplicity of MD
 152 compared to other markup languages. **Fig. 3** illustrates the generation of various formatted documents
 153 from a manuscript in pandoc MD. Some relevant functions for scientific texts are explained below in
 154 more detail.

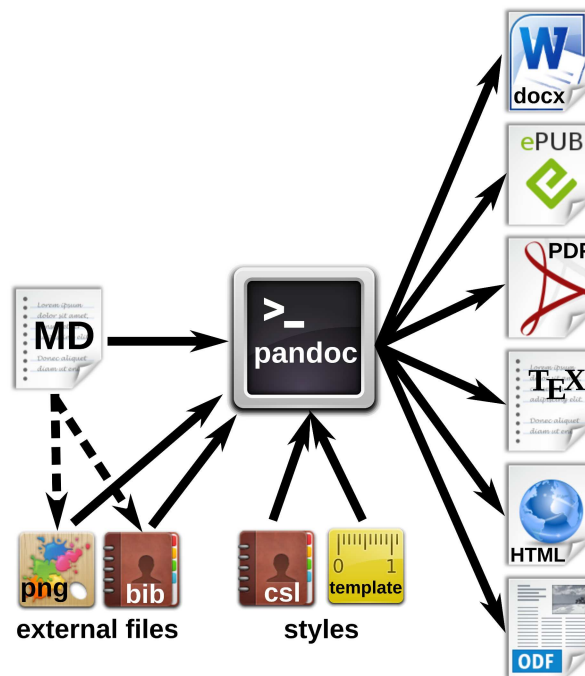


Figure 3. Workflow for the generation of multiple document formats with pandoc. The markdown (MD) file contains the manuscript text with formatting tags, and can also refer to external files such as images or reference databases. The pandoc processor converts the MD file to the desired output formats. Documents, citations etc. can be defined in style files or templates.

155 MARKDOWN EDITORS AND ONLINE EDITING

156 The usability of a text editor is important for the author, either writing alone or with several co-authors. In
 157 this section we present software and strategies for different scenarios. **Fig. 4** summarizes various options
 158 for local or networked editing of MD files.

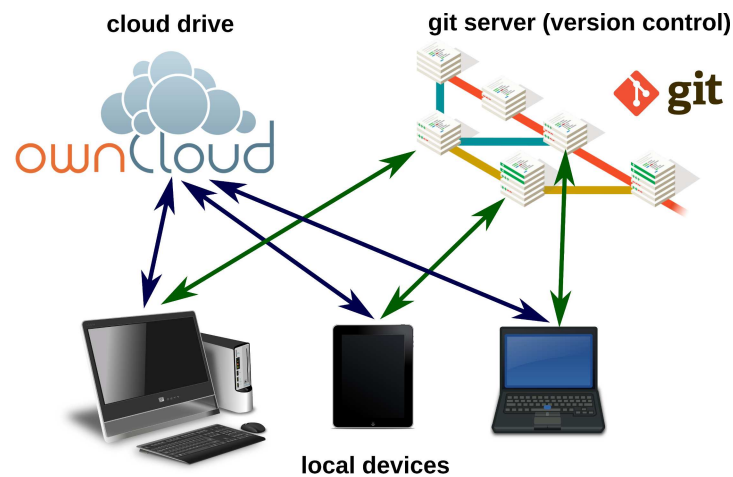


Figure 4. Markdown files can be edited on local devices or on cloud drives. A local or remote git repository enables advanced advanced version control.

159 Markdown editors

160 Due to MD's simple syntax, basically any text editor is suitable for editing markdown files. The formatting
 161 tags are written in plain text and are easy to remember. Therefore, the author is not distracted by looking
 162 around for layout options with the mouse. For several popular text editors, such as vim (<http://www.vim.org/>), GNU Emacs (<https://www.gnu.org/software/emacs/>), atom (<https://atom.io/>)
 163 or geany (<http://www.geany.org/>), plugins provide additional functionality for markdown editing,
 164 e.g. syntax highlighting, command helpers, live preview or structure browsing.
 165

166 Various dedicated markdown editors have been published as well. Many of those are cross-platform com-
 167 patible, such as Abricotine (<http://abricotine.brrd.fr/>), ghostwriter (<https://github.com/wereturtle/ghostwriter>) and CuteMarkEd (<https://cloose.github.io/CuteMarkEd/>).
 168

169 The lightweight format is also ideal for writing on mobile devices. Numerous applications are available on
 170 the App stores for Android and iOS systems. The programs Swype and Dragon (<http://www.nuance.com/>)
 171 facilitate the input of text on such devices by guessing words from gestures and speech recognition
 172 (dictation).

173 **Fig. 5.** shows the editing of a markdown file, using the cross-platform editor Atom with several markdown
 174 plugins.

175 Online editing and collaborative writing

176 Storing manuscripts on network drives (*The Cloud*) has become popular for several reasons:

- 177 • Protection against data loss.
- 178 • Synchronization of documents between several devices.
- 179 • Collaborative editing options.

180 Markdown files on a Google Drive (<https://drive.google.com>) for instance can be edited online
 181 with StackEdit (<https://stackedit.io>). **Fig. 6** demonstrates the online editing of a markdown file
 182 on an ownCloud (<https://owncloud.com/>) installation. OwnCloud is an Open Source software plat-
 183 form, which allows the set-up of a file server on personal webspace. The functionality of an ownCloud
 184 installation can be enhanced by installing plugins.

185 Even mathematical formulas are rendered correctly in the HTML live preview window of the ownCloud
 186 markdown plugin (**Fig. 6**).

187 The collaboration and authoring platform Authorea (<https://www.authorea.com/>) also supports
 188 markdown as one of multiple possible input formats. This can be beneficial for collaborations in which
 189 one or more authors are not familiar with markdown syntax.

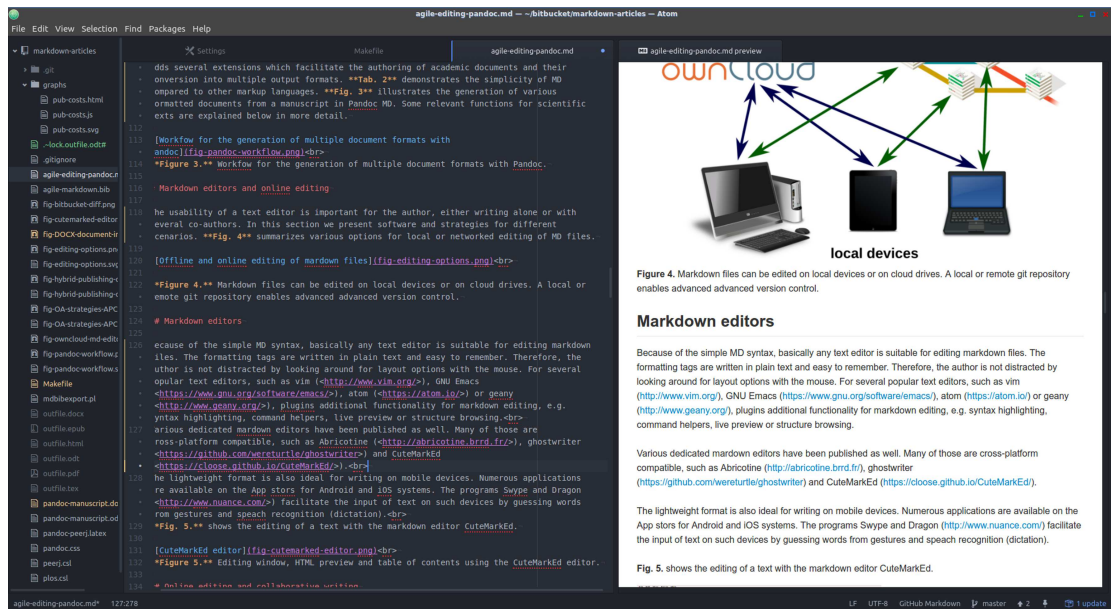


Figure 5. Document directory tree, editing window and HTML preview using the Atom editor.

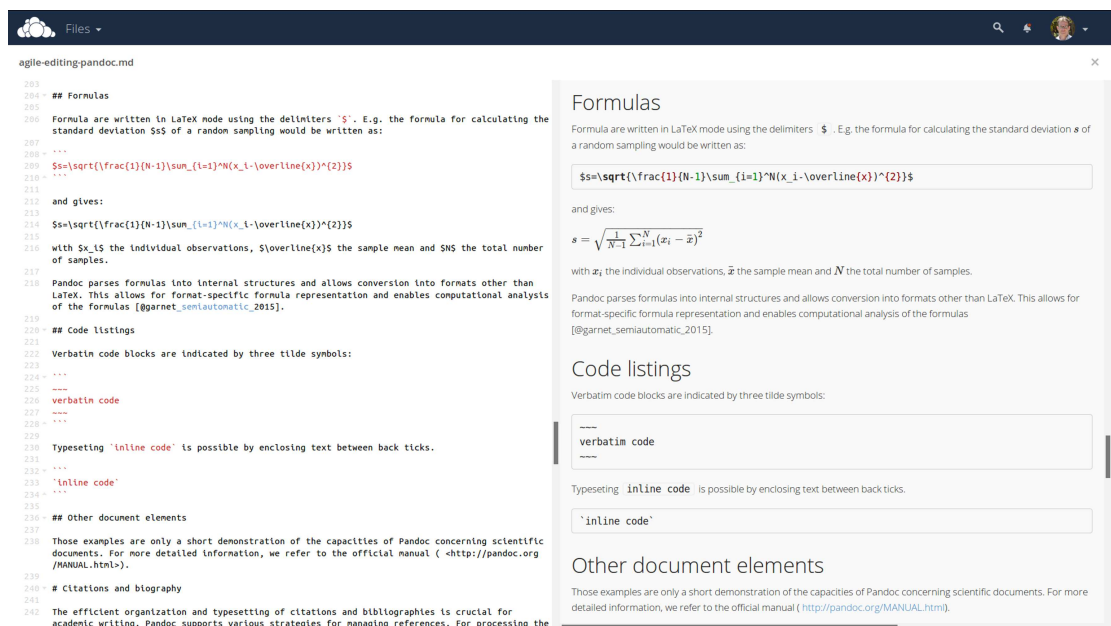


Figure 6. Direct online editing of this manuscript with live preview using the ownCloud Markdown Editor plugin by Robin Appelman.

190 Document versioning and change control

191 Programmers, especially when working in distributed teams, rely on version control systems to manage
 192 changes of code. Currently, Git (<https://git-scm.com/>), which is also used e.g. for the development
 193 of the Linux kernel, is one of the most employed software solutions for versioning. Git allows the parallel
 194 work of collaborators and has an efficient merging and conflict resolution system. A Git repository may
 195 be used by a single local author to keep track of changes, or by a team with a remote repository, e.g. on
 196 github (<https://github.com/>) or bitbucket (<https://bitbucket.org/>). Because of the plain text
 197 format of markdown, Git can be used for version control and distributed writing. For the writing of the
 198 present article, the co-authors (Germany and Mexico) used a remote Git repository on bitbucket. The
 199 plain text syntax of markdown facilitates the visualization of differences of document versions, as shown
 200 in **Fig. 7**.

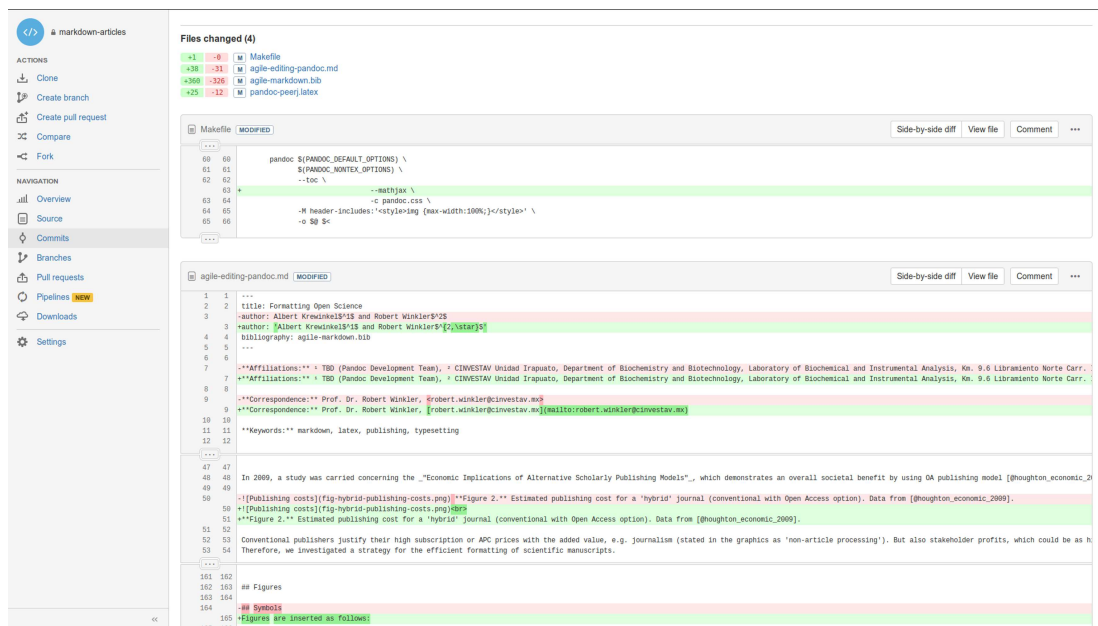


Figure 7. Version control and collaborative editing using a git repository on bitbucket.

201 PANDOC MARKDOWN FOR SCIENTIFIC TEXTS

202 In the following section, we demonstrate the potential for typesetting scientific manuscripts with pan-
 203 doc using examples for typical document elements, such as tables, figures, formulas, code listings and
 204 references. A brief introduction is given by Dominici (2014). The complete Pandoc User's Manual is
 205 available at <http://pandoc.org/MANUAL.html>.

206 Tables

207 There are several options to write tables in markdown. The most flexible alternative - which was also
 208 used for this article - are pipe tables. The contents of different cells are separated by pipe symbols (|):

```
209 Left | Center | Right | Default
210 :----|:-----|:-----|:-----
211 LLL | CCC   | RRR   | DDD
212 gives
```

Left	Center	Right	Default
LLL	CCC	RRR	DDD

213 The headings and the alignment of the cells are given in the first two lines. The cell width is variable. The
 214 pandoc parameter `--columns=NUM` can be used to define the length of lines in characters. If contents do
 215 not fit, they will be wrapped.

216 Complex tables, e.g. tables featuring multiple headers or those containing cells spanning multiple rows or
 217 columns, are currently not representable in markdown format. However, it is possible to embed LATEX
 218 and HTML tables into the document. These format-specific tables will only be included in the output if
 219 a document of the respective format is produced. This is method can be extended to apply any kind of
 220 format-specific typographic functionality which would otherwise be unavailable in markdown syntax.

221 **Figures and images**

222 Images are inserted as follows:

223 `![alt text](image location/ name)`

224 e.g.

225 `![Publishing costs](fig-hybrid-publishing-costs.png)`

226 The *alt text* is used e.g. in HTML output. Image dimensions can be defined in braces:

227 ``

228 As well, an identifier for the figure can be defined with #, resulting e.g. in the image attributes `{#figure1`
 229 `height=30%}`.

230 A paragraph containing only an image is interpreted as a figure. The *alt text* is then output as the figure's
 231 caption.

232 **Symbols**

233 Scientific texts often require special characters, e.g. Greek letters, mathematical and physical symbols
 234 etc.

235 The UTF-8 standard, developed and maintained by *Unicode Consortium*, enables the use of characters
 236 across languages and computer platforms. The encoding is defined as RFC document 3629 of the Network
 237 Working group (Yergeau, 2003) and as ISO standard ISO/IEC 10646:2014 (International Organization for
 238 Standardization, 2014). Specifications of Unicode and code charts are provided on the Unicode homepage
 239 (<http://www.unicode.org/>).

240 In pandoc markdown documents, Unicode characters such as °, α, ä, Å can be inserted directly and
 241 passed to the different output documents. The correct processing of MD with UTF-8 encoding to LA-
 242 TEX/PDF output requires the use of the `--latex-engine=xelatex` option and the use of an appropriate
 243 font. The Times-like XITS font (<https://github.com/khaledhosny/xits-math>), suitable for high
 244 quality typesetting of scientific texts, can be set in the LATEX template:

```
\usepackage{unicode-math}
\setmainfont
[
  Extension = .otf,
  UprightFont = *-regular,
  BoldFont = *-bold,
  ItalicFont = *-italic,
  BoldItalicFont = *-bolditalic,
]{xits}
\setmathfont
[
  Extension = .otf,
  BoldFont = *bold,
]{xits-math}
```

245 To facilitate the input of specific characters, so-called mnemonics can be enabled in some editors (e.g. in
 246 atom by the `character-table` package). For example, the 2-character Mnemonics ‘:u’ gives ‘ü’ (di-
 247 aeresis), or ‘D*’ the Greek Δ. The possible character mnemonics and character sets are listed in RFC
 248 1345 <http://www.faqs.org/rfcs/rfc1345.html> (Simonsen, 1992).

249 Formulas

250 Formulas are written in LATEX mode using the delimiters \$. E.g. the formula for calculating the standard
 251 deviation s of a random sampling would be written as:

$$252 \text{\$s}=\sqrt{\frac{1}{N-1}\sum_{i=1}^N(x_i-\overline{x})^2}\text{\$}$$

253 and gives:

$$254 s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

255 with x_i the individual observations, \bar{x} the sample mean and N the total number of samples.

256 Pandoc parses formulas into internal structures and allows conversion into formats other than LATEX.
 257 This allows for format-specific formula representation and enables computational analysis of the formulas
 258 (Corbí & Burgos, 2015).

259 Code listings

260 Verbatim code blocks are indicated by three tilde symbols:

```
261 ~~~
262 verbatim code
263 ~~~
```

264 Typesetting `inline` code is possible by enclosing text between back ticks.

```
265 `inline code`
```

266 Other document elements

267 These examples are only a short demonstration of the capacities of pandoc concerning scientific docu-
 268 ments. For more detailed information, we refer to the official manual (<http://pandoc.org/MANUAL.html>).
 269 [html](http://pandoc.org/MANUAL.html)).

270 CITATIONS AND BIOGRAPHY

271 The efficient organization and typesetting of citations and bibliographies is crucial for academic writing.
 272 Pandoc supports various strategies for managing references. For processing the citations and the creation
 273 of the bibliography, the command line parameter `--filter pandoc-citeproc` is used, with variables
 274 for the reference database and the bibliography style. The bibliography will be located automatically at
 275 the header `# References` or `# Bibliography`.

276 Reference databases

277 Pandoc is able to process all mainstream literature database formats, such as RIS, BIB, etc. However, for
 278 maintaining compatibility with LATEX/ BIBTEX, the use of BIB databases is recommended. The used
 279 database either can be defined in the YAML metablock of the MD file (see below) or it can be passed as
 280 parameter when calling pandoc.

281 **Inserting citations**

282 For inserting a reference, the database key is given within square brackets, and indicated by an '@'. It is
283 also possible to add information, such as page:

```
284 [@suber_open_2012; @benkler_wealth_2006, 57 ff.]
```

285 gives (Benkler, 2006, p. 57 ff.; Suber, 2012).

286 **Styles**

287 The Citation Style Language (CSL) <http://citationstyles.org/> is used for the citations and bibli-
288 ographies. This file format is supported e.g. by the reference management programs Mendeley <https://www.mendeley.com/>,
289 Papers <http://papersapp.com/> and Zotero <https://www.zotero.org/>.
290 CSL styles for particular journals can be found from the Zotero style repository [https://www.zotero.org/](https://www.zotero.org/styles)
291 [org/styles](https://www.zotero.org/styles). The bibliography style that pandoc should use for the target document can be chosen in
292 the YAML block of the markdown document or can be passed in as an command line option. The latter
293 is more recommendable, because distinct bibliography style may be used for different documents.

294 **Creation of LATEX natbib citations**

295 For citations in scientific manuscripts written in LATEX, the natbib package is widely used. To create
296 a LATEX output file with natbib citations, pandoc simply has to be run with the `--natbib` option, but
297 without the `--filter pandoc-citeproc` parameter.

298 **Database of cited references**

299 To share the bibliography for a certain manuscript with co-authors or the publisher's production team, it
300 is often desirable to generate a subset of a larger database, which only contains the cited references. If
301 LATEX output was generated with the `--natbib` option, the compilation of the file with LATEX gives an
302 AUX file (in the example named `md-article.aux`), which subsequently can be extracted using BibTool
303 <https://github.com/ge-ne/bibtool>:

```
304 ~~~  
305 bibtool -x md-article.aux -o bibshort.bib  
306 ~~~
```

307 In this example, the article database will be called `bibshort.bib`.

308 For the direct creation of an article specific BIB database without using LATEX, we wrote a simple Perl
309 script called `mdbibexport` (<https://github.com/robert-winkler/mdbibexport>).

310 **META INFORMATION OF THE DOCUMENT**

311 Bourne (2005) argues that journals should be effectively equivalent to biological databases: both provide
312 data which can be referenced by unique identifiers like DOI or e.g. gene IDs. Applying the semantic-web
313 ideas of Berners-Lee & Hendler (2001) to this domain can make this vision a reality. Here we show how
314 metadata can be specified in markdown. We propose conventions, and demonstrate their suitability to
315 enable interlinked and semantically enriched journal articles.

316 Document information such as title, authors, abstract etc. can be defined in a metadata block written in
317 YAML syntax. YAML ("YAML Ain't Markup Language", <http://yaml.org/>) is a data serialization
318 standard in simple, human readable format. Variables defined in the YAML section are processed by
319 pandoc and integrated into the generated documents. The YAML metadata block is recognized by three
320 hyphens (`---`) at the beginning, and three hyphens or dots (`...`) at the end, e.g.:

```

---
title: Formatting Open Science
subtitle: agile creation of multiple document types
date: 2017-02-10
...

```

321 The public availability of all relevant information is a central aspect of Open Science. Analogous to article
 322 contents, data should be accessible via default tools. We believe that this principle must also be applied
 323 to article metadata. Thus, we created a custom pandoc writer that emits the article's data as JSON-LD
 324 (Lanthaler & Gütl, 2012), allowing for informational and navigational queries of the journal's data with
 325 standard tools of the semantic web. The above YAML information would be output as:

```

{
  "@context": {
    "@vocab": "http://schema.org/",
    "date": "datePublished",
    "title": "headline",
    "subtitle": "alternativeTitle"
  },
  "@type": "ScholarlyArticle",
  "title": "Formatting Open Science",
  "subtitle": "agile creation of multiple document types",
  "date": "2017-02-10"
}

```

326 This format allows processing of the information by standard data processing software and browsers.

327 Flexible metadata authoring

328 We developed a method to allow writers the flexible specification of authors and their respective affili-
 329 ations. Author names can be given as a string, via the key of a single-element object, or explicitly as a
 330 name attribute of an object. Affiliations can be specified directly as properties of the author object, or
 331 separately in the institute object.

332 Additional information, e.g. email addresses or identifiers like ORCID (Haak et al., 2012), can be added
 333 as additional values:

```

author:
  - John Doe:
      institute: fs
      email: john.doe@example.com
      orcid: 0000-0000-0000-0000
institute:
  fs: Science Formatting Working Group

```

334 JATS support

335 The journal article tag suite (JATS) was developed by the NLM and standardized by ANSI/NISO as
 336 an archiving and exchange format of journal articles and the associated metadata (National Information
 337 Standards Organization, 2012), including data of the type shown above. The `pandoc-jats` writer by
 338 Martin Fenner is a plugin usable with pandoc to produce JATS-formatted output. The writer was adapted
 339 to be compatible with our metadata authoring method, allowing for simple generation of files which
 340 contain the relevant metadata.

341 Citation types

342 Writers can add information about the reason a citation is given. This might help reviewers and readers,
 343 and can simplify the search for relevant literature. We developed an extended citation syntax that inte-

344 grates seamlessly into markdown and can be used to add complementary information to citations. Our
 345 method is based on CiTO, the Citation Typing Ontology (Shotton, 2010), which specifies a vocabulary
 346 for the motivation when citing a resource. The type of a citations can be added to a markdown citation us-
 347 ing @CITO_PROPERTY:KEY, where CITO_PROPERTY is a supported CiTO property, and KEY is the usual
 348 citation key. Our tool extracts that information and includes it in the generated linked data output. A
 349 general CiTO property (*cites*) is used, if no CiTO property is found in a citation key.

350 The work at hand will always be the subject of the generated semantic *subject-predicate-object* triples.
 351 Some CiTO predicates cannot be used in a sensical way under this condition. Focusing on author conve-
 352 nience, we use this fact to allow shortening of properties when sensible. E.g. if authors of a biological
 353 paper include a reference to the paper describing a method which was used in their work, this relation
 354 can be described by the *uses_method_in* property of the CiTO ontology. The inverse property, *pro-
 355 vides_method_for*, would always be nonsensical in this context as implied by causality. It is therefor not
 356 supported by our tool. This allows us to introduce an abbreviation (*method*) for the latter property, as any
 357 ambiguity has been eliminated. Users of western blotting might hence write @method_in:towbin_1979
 358 or even just @method:towbin_1979, where *towbin_1979* is the citation identifier of the describing paper
 359 by Towbin, Staehelin & Gordon (1979).

360 **EXAMPLE: MANUSCRIPT WITH OUTPUT OF DOCX/ ODT FORMAT** 361 **AND LATEX/ PDF FOR SUBMISSION TO DIFFERENT JOURNALS.**

362 Scientific manuscripts have to be submitted in a format defined by the journal or publisher. At the moment,
 363 DOCX is the most common file format for manuscript submission. Some publishers also accept or require
 364 LATEX or ODT formats. Additional to the general style of the manuscript - organization of sections,
 365 fonts, etc. – the citation style of the journal must also be followed. Often, the same manuscript has to be
 366 prepared for different journals, e.g. if the manuscript was rejected by a journal and has to be formatted
 367 for another one, or if a preprint of the paper is submitted to an archive that requires a distinct document
 368 format than the targeted peer-reviewed journal. In this example, we want to create a manuscript for a
 369 *PLoS* journal in DOCX and ODT format for WYSIWYG word processors. Further, a version in LATEX/
 370 PDF should be produced for PeerJ submission and archiving at the PeerJ preprint server.

371 The examples for DOCX/ ODT are kept relatively simple, to show the proof-of-principle and to provide a
 372 plain document for the development of own templates. Nevertheless, the generated documents should be
 373 suitable for submission after little manual editing. For specific journals it may be necessary to create more
 374 sophisticated templates or to copy/ paste the generic DOCX/ ODT output into the publisher's template.

375 **Development of a DOCX/ ODT template**

376 A first DOCX document with bibliography in *PLoS* format is created with pandoc DOCX output:

```
377 pandoc -S -s --csl=plos.csl --filter pandoc-citeproc  
378 -o pandoc-manuscript.docx agile-editing-pandoc.md
```

379 The parameters `-S -s` generate a typographically correct (dashes, non-breaking spaces etc.) stand-alone
 380 document. A bibliography with the *PLoS* style is created by the citeproc filter setting `--csl=plos.csl`
 381 `--filter pandoc-citeproc`.

382 The document settings and styles of the resulting file `pandoc-manuscript.docx` can be optimized and
 383 be used again as document template (`--reference-docx=pandoc-manuscript.docx`).

```
384 pandoc -S -s --reference-docx=pandoc-manuscript.docx --csl=plos.csl  
385 --filter pandoc-citeproc -o outfile.docx agile-editing-pandoc.md
```

386 It is also possible to directly re-use a previous output file as template (i.e. template and output file have
 387 the same file name):

```
388 pandoc -S -s --columns=10 --reference-docx=pandoc-manuscript.docx  
389 --csl=plos.csl --filter=pandoc-citeproc  
390 -o pandoc-manuscript.docx agile-editing-pandoc.md
```


384 In this way, the template can be incrementally adjusted to the desired document formatting. The final
 385 document may be employed later as pandoc template for other manuscripts with the same specifications.
 386 In this case, running pandoc the first time with the template, the contents of the new manuscript would
 387 be filled into the provided DOCX template. A page with DOCX manuscript formatting of this article is
 388 shown in **Fig. 8**.

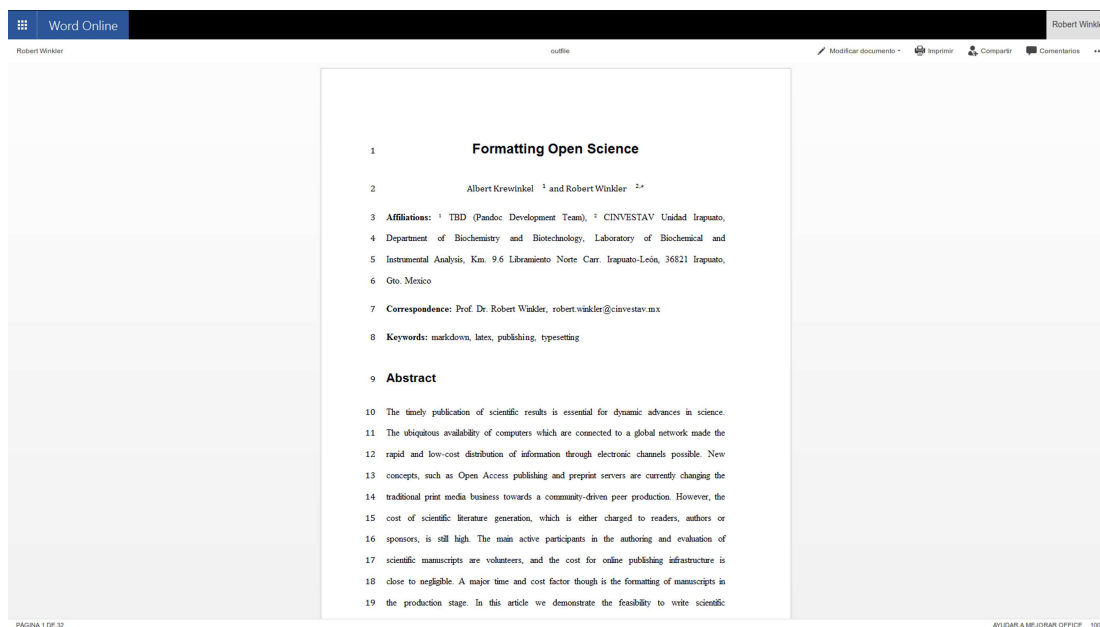


Figure 8. Opening a pandoc-generated DOCX in Microsoft Office 365.

389 The same procedure can be applied with an ODT formatted document.

390 Development of a TEX/PDF template

391 The default pandoc LATEX template can be written into a separate file by:

```
392 pandoc -D latex > template-peerj.tex
```

393 This template can be adjusted, e.g. by defining Unicode encoding (see above), by including particular
 394 packages or setting document options (line numbering, font size). The template can then be used with
 the pandoc parameter `--template=pandoc-peerj.tex`.

395 The templates used for this document are included as Supplemental Material (see section *Software and*
 396 *code availability* below).

397 Styles for HTML and EPUB

398 The style for HTML and EPUB formats can be defined in .css stylesheets. The Supplemental Material
 399 contains a simple example .css file for modifying the HTML output, which can be used with the pandoc
 400 parameter `-c pandoc.css`.

401 AUTOMATING DOCUMENT PRODUCTION

402 The commands necessary to produce the document in a specific formats or styles can be defined in a
 403 simple Makefile. An example Makefile is included in the source code of this preprint. The desired
 404 output file format can be chosen when calling make. E.g. `make outfile.pdf` produces this preprint in
 405 PDF format. Calling `make` without any option creates all listed document types. A Makefile producing
 406 DOCX, ODT, JATS, PDF, LATEX, HTML and EPUB files of this document is provided as Supplemental
 407 Material.

408 **Cross-platform compatibility**

409 The make process was tested on Windows 10 and Linux 64 bit. All documents – DOCX, ODT, JATS,
410 LATEX, PDF, EPUB and HTML – were generated successfully, which demonstrates the cross-platform
411 compatibility of the workflow.

412 **PERSPECTIVE**

413 Following the trend to peer production, the formatting of scientific content must become more efficient.
414 Markdown/ pandoc has the potential to play a key role in the transition from proprietary to community-
415 driven academic production. Important research tools, such as the statistical computing and graph-
416 ics language R (R Core Team, 2014) and the Jupyter notebook project (Kluyver et al., 2016) have al-
417 ready adopted the MD syntax (e.g. <http://rmarkdown.rstudio.com/>). The software for writing
418 manuscripts in MD is mature enough to be used by academic writers. Therefore, publishers also should
419 consider implementing the MD format into their editorial platforms.

420 **CONCLUSIONS**

421 Authoring scientific manuscripts in markdown (MD) format is straight-forward, and manual formatting is
422 reduced to a minimum. The simple syntax of MD facilitates document editing and collaborative writing.
423 The rapid conversion of MD to multiple formats such as DOCX, LATEX, PDF, EPUB and HTML can
424 be done easily using pandoc, and templates enable the automated generation of documents according to
425 specific journal styles.

426 The additional features we implemented facilitate the correct indexing of meta information of journal
427 articles according to the ‘semantic web’ philosophy.

428 Altogether, the MD format supports the agile writing and fast production of scientific literature. The
429 associated time and cost reduction especially favours community-driven publication strategies.

430 **ACKNOWLEDGMENTS**

431 We cordially thank Dr. Gerd Neugebauer for his help in creating a subset of a bibtex data base using
432 BibTool, as well as Dr. Ricardo A. Chávez Montes, Prof. Magnus Palmblad and Martin Fenner for com-
433 ments on the manuscript. Warm thanks also go to Anubhav Kumar and Jennifer König for proofreading.
434 The work was funded by the Consejo Nacional de Ciencia y Tecnología (CONACyT) Mexico, with the
435 grant FRONTERAS 2015-2/814 and by institutional funding of the Centro de Investigación y de Estudios
436 Avanzados del Instituto Politécnico Nacional (CINVESTAV).

437 **SOFTWARE AND CODE AVAILABILITY**

438 The relevant software for creating this manuscript used is cited according to (Smith, Katz & Niemeyer,
 439 2016) and listed in **Tab. 3**. Since unique identifiers are missing for most software projects, we only refer
 440 to the project homepages or software repositories:

Table 4. Relevant software used for this article.

Software	Use	Authors	Version	Release	Homepage/ repository
pandoc	universal markup converter	John MacFarlane	1.16.0.2	16/01/13	http://www.pandoc.org
pandoc-citeproc	library for CSL citations with pandoc	John MacFarlane, Andrea Rossato	0.9.1	16/03/19	https://github.com/jgm/pandoc-citeproc
pandoc-jats	creation of JATS files with pandoc	Martin Fenner	0.9	15/04/26	https://github.com/mfenner/pandoc-jats
ownCloud	personal cloud software	ownCloud GmbH, Community	9.1.1	16/09/20	https://owncloud.org/
Markdown Editor	plugin for ownCloud	Robin Appelman	0.1	16/03/08	https://github.com/icewind1991/files_markdown
BibTool	Bibtex database tool	Gerd Neugebauer	2.63	16/01/16	https://github.com/ge-ne/bibttool

441 The software created as part of this article, *pandoc-scholar*, is suitable for general use and has been published at <https://github.com/pandoc-scholar/pandoc-scholar>, DOI: 10.5281/zenodo.376761.
 442
 443 The source code of this manuscript, as well as the templates and pandoc Makefile, have been deposited to <https://github.com/robert-winkler/scientific-articles-markdown/>.
 444

445 Drawings for document types, devices and applications have been adopted from Calibre <http://calibre-ebook.com/>, openclipart <https://openclipart.org/> and the GNOME Theme Faenza <https://code.google.com/archive/p/faenza-icon-theme/>.
 446
 447

448 **BIBLIOGRAPHY**

- 449 Benkler Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*.
450 New Haven, CT, USA: Yale University Press.
- 451 Berners-Lee T., Hendler J. 2001. Publishing on the semantic web. *Nature* 410:1023–1024. DOI:
452 10.1038/35074206.
- 453 Bourne P. 2005. Will a biological database be different from a biological journal? *PLOS Computational*
454 *Biology* 1:e34. DOI: 10.1371/journal.pcbi.0010034.
- 455 Brauer M., Durusau P., Edwards G., Faure D., Magliery T., Vogelheim D. 2005. *Open Document Format*
456 *for Office Applications (OpenDocument) v1.0*. OASIS.
- 457 Brown C. 2001. The E-Volution of Preprints in the Scholarly Communication of Physicists and As-
458 tronomers. *J. Am. Soc. Inf. Sci.* 52:187–200. DOI: 10.1002/1097-4571(2000)9999:9999<:AID-
459 ASI1586>3.0.CO;2-D.
- 460 Brown C. 2003. The Role of Electronic Preprints in Chemical Communication: Analysis of Cita-
461 tion, Usage, and Acceptance in the Journal Literature. *J. Am. Soc. Inf. Sci.* 54:362–371. DOI:
462 10.1002/asi.10223.
- 463 Brown PO., Eisen MB., Varmus HE. 2003. Why PLoS Became a Publisher. *PLoS Biol* 1. DOI:
464 10.1371/journal.pbio.0000036.
- 465 Butler D. 2001. Los Alamos Loses Physics Archive as Preprint Pioneer Heads East. *Nature* 412:3–4.
466 DOI: 10.1038/35083708.
- 467 Callaway E. 2013. Preprints Come to Life. *Nature News* 503:180. DOI: 10.1038/503180a.
- 468 Corbí A., Burgos D. 2015. Semi-Automated Correction Tools for Mathematics-Based Exercises in
469 MOOC Environments. *International Journal of Interactive Multimedia and Artificial Intelligence* 3:89–
470 95. DOI: 10.9781/ijimai.2015.3312.
- 471 Dominici M. 2014. An overview of Pandoc. *TUGboat* 35:44–50.
- 472 DPT Collective. 2015. From Print to Ebooks: A Hybrid Publishing Toolkit for the Arts. In: Monk J,
473 Rasch M, Cramer F, Wu A eds. Institute of Network Cultures,
- 474 Eikebrokk T., Dahl TA., Kessel S. 2014. EPUB as Publication Format in Open Access Journals: Tools
475 and Workflow. *Code4Lib*.
- 476 Eisen M. 2003. Publish and be praised. *The Guardian*.
- 477 Fecher B., Friesike S. 2014. Open Science: One Term, Five Schools of Thought. In: Bartling S, Friesike
478 S eds. *Opening Science*. Springer International Publishing, 17–47.
- 479 Ginsparg P. 1994. First Steps Towards Electronic Research Communication. *Computers in Physics*
480 8:390–396. DOI: 10.1063/1.4823313.
- 481 Haak LL., Fenner M., Paglione L., Pentz E., Ratner H. 2012. ORCID: A system to uniquely identify
482 researchers. *Learned Publishing* 25:259–264. DOI: 10.1087/20120404.
- 483 Hickson I., Berjon R., Faulkner S., Leithead T., Navara ED., O’Connor E., Pfeiffer S., Faulkner S., Navara
484 ED., Leithead T., Berjon R., Hickson I., Pfeiffer S., O’Connor T. 2014. *HTML5*. W3C.
- 485 Houghton J., Rasmussen B., Sheehan P., Oppenheim C., Morris A., Creaser C., Greenwood H., Summers
486 M., Gourlay A. 2009. Economic implications of alternative scholarly publishing models: Exploring the
487 costs and benefits.
- 488 International Organization for Standardization. 2013. ISO 32000-1:2008 - Document management –
489 Portable document format – Part 1: PDF 1.7. *ISO*.
- 490 International Organization for Standardization. 2014. ISO/IEC 10646:2014 - Information technology –

- 491 Universal Coded Character Set (UCS). *ISO*.
- 492 Kielhorn A. 2011. Multi-target publishing-Generating ePub, PDF, and more, from Markdown using
493 pandoc. *TUGboat-TeX Users Group* 32:272.
- 494 Kluyver T., Ragan-Kelley B., Pérez F., Granger B., Bussonnier M., Frederic J., Kelley K., Hamrick J.,
495 Grout J., Corlay S., others. 2016. Jupyter notebooks—a publishing format for reproducible computational
496 workflows. In: *Positioning and power in academic publishing: Players, agents and agendas*. 87–90.
497 DOI: 10.3233/978-1-61499-649-1-87.
- 498 Lamport L. 1994. *LaTeX: A Document Preparation System*. Reading, Mass: Addison-Wesley Profes-
499 sional.
- 500 Lanthaler M., Gütl C. 2012. On using JSON-LD to create evolvable RESTful services. In: *Proceedings*
501 *of the third international workshop on RESTful design*. ACM, 25–32.
- 502 Leonard S. 2016. *Guidance on Markdown: Design Philosophies, Stability Strategies, and Select Regis-*
503 *trations*. RFC Editor; Internet Request for Comments.
- 504 National Information Standards Organization. 2012. *JATS: Journal Article Tag Suite*.
- 505 Ngo T. 2006. *OFFICE OPEN XML OVERVIEW ECMA TC45*. Ecma International.
- 506 Ovidia S. 2014. Markdown for Librarians and Academics. *Behavioral & Social Sciences Librarian*
507 33:120–124. DOI: 10.1080/01639269.2014.904696.
- 508 R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna, Austria: R
509 Foundation for Statistical Computing.
- 510 Raggett D., Hors AL., Jacobs I., Le Hors A., Raggett D., Jacobs I. 1999. *HTML 4.01 Specification*. W3C.
- 511 Shotton D. 2010. CiTO, the Citation Typing Ontology. *Journal of Biomedical Semantics* 1:S6. DOI:
512 10.1186/2041-1480-1-S1-S6.
- 513 Simonsen K. 1992. *Character Mnemonics & Character Sets*. Rationel Almen Planlaegning; Internet
514 Request for Comments.
- 515 Smith AM., Katz DS., Niemeyer KE. 2016. Software Citation Principles. *PeerJ Computer Science* 2:e86.
516 DOI: 10.7717/peerj-cs.86.
- 517 Solomon D., Björk B-C. 2016. Article Processing Charges for Open Access Publicationthe Situation for
518 Research Intensive Universities in the USA and Canada. *PeerJ* 4:e2264. DOI: 10.7717/peerj.2264.
- 519 Suber P. 2012. *Open Access*. Cambridge, Mass: The MIT Press.
- 520 Towbin H., Staehelin T., Gordon J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to
521 nitrocellulose sheets: Procedure and some applications. *Proceedings of the National Academy of Sciences*
522 76:4350–4354.
- 523 Van Noorden R. 2012. Journal Offers Flat Fee for “all You Can Publish”. *Nature News* 486:166. DOI:
524 10.1038/486166a.
- 525 Van Noorden R. 2013. Open Access: The True Cost of Science Publishing. *Nature* 495:426–429. DOI:
526 10.1038/495426a.
- 527 Van Noorden R. 2014. The arXiv Preprint Server Hits 1 Million Articles. *Nature News*. DOI: 10.1038/na-
528 ture.2014.16643.
- 529 Volmer DA., Stokes CS. 2016. How to Prepare a Manuscript Fit-for-Purpose for Submission and Avoid
530 Getting a “desk-Reject”. *Rapid Commun. Mass Spectrom.*:n/a–n/a. DOI: 10.1002/rcm.7746.
- 531 Willinsky J. 2005. The Unacknowledged Convergence of Open Source, Open Access, and Open Science.
532 *First Monday* 10. DOI: 10.5210/fm.v10i8.1265.
- 533 Woelfle M., Olliaro P., Todd MH. 2011. Open Science Is a Research Accelerator. *Nat Chem* 3:745–748.

- 534 DOI: 10.1038/nchem.1149.
- 535 Yergeau F. 2003. *UTF-8, a transformation format of ISO 10646*. Alis Technologies.
- 536 Youngen GK. 1998. Citation Patterns to Traditional and Electronic Preprints in the Published Literature.
537 *Coll. res. libr.* 59:448–456. DOI: 10.5860/crl.59.5.448.
- 538 Benkler Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*.
539 New Haven, CT, USA: Yale University Press.
- 540 Berners-Lee T., Hendler J. 2001. Publishing on the semantic web. *Nature* 410:1023–1024. DOI:
541 10.1038/35074206.
- 542 Bourne P. 2005. Will a biological database be different from a biological journal? *PLOS Computational*
543 *Biology* 1:e34. DOI: 10.1371/journal.pcbi.0010034.
- 544 Brauer M., Durusau P., Edwards G., Faure D., Magliery T., Vogelheim D. 2005. *Open Document Format*
545 *for Office Applications (OpenDocument) v1.0*. OASIS.
- 546 Brown C. 2001. The E-Volution of Preprints in the Scholarly Communication of Physicists and As-
547 tronomers. *J. Am. Soc. Inf. Sci.* 52:187–200. DOI: 10.1002/1097-4571(2000)9999:9999<:AID-
548 ASI1586>3.0.CO;2-D.
- 549 Brown C. 2003. The Role of Electronic Preprints in Chemical Communication: Analysis of Cita-
550 tion, Usage, and Acceptance in the Journal Literature. *J. Am. Soc. Inf. Sci.* 54:362–371. DOI:
551 10.1002/asi.10223.
- 552 Brown PO., Eisen MB., Varmus HE. 2003. Why PLoS Became a Publisher. *PLoS Biol* 1. DOI:
553 10.1371/journal.pbio.0000036.
- 554 Butler D. 2001. Los Alamos Loses Physics Archive as Preprint Pioneer Heads East. *Nature* 412:3–4.
555 DOI: 10.1038/35083708.
- 556 Callaway E. 2013. Preprints Come to Life. *Nature News* 503:180. DOI: 10.1038/503180a.
- 557 Corbí A., Burgos D. 2015. Semi-Automated Correction Tools for Mathematics-Based Exercises in
558 MOOC Environments. *International Journal of Interactive Multimedia and Artificial Intelligence* 3:89–
559 95. DOI: 10.9781/ijimai.2015.3312.
- 560 Dominici M. 2014. An overview of Pandoc. *TUGboat* 35:44–50.
- 561 DPT Collective. 2015. From Print to Ebooks: A Hybrid Publishing Toolkit for the Arts. In: Monk J,
562 Rasch M, Cramer F, Wu A eds. Institute of Network Cultures,
- 563 Eikebrokk T., Dahl TA., Kessel S. 2014. EPUB as Publication Format in Open Access Journals: Tools
564 and Workflow. *Code4Lib*.
- 565 Eisen M. 2003. Publish and be praised. *The Guardian*.
- 566 Fecher B., Friesike S. 2014. Open Science: One Term, Five Schools of Thought. In: Bartling S, Friesike
567 S eds. *Opening Science*. Springer International Publishing, 17–47.
- 568 Ginsparg P. 1994. First Steps Towards Electronic Research Communication. *Computers in Physics*
569 8:390–396. DOI: 10.1063/1.4823313.
- 570 Haak LL., Fenner M., Paglione L., Pentz E., Ratner H. 2012. ORCID: A system to uniquely identify
571 researchers. *Learned Publishing* 25:259–264. DOI: 10.1087/20120404.
- 572 Hickson I., Berjon R., Faulkner S., Leithead T., Navara ED., O’Connor E., Pfeiffer S., Faulkner S., Navara
573 ED., Leithead T., Berjon R., Hickson I., Pfeiffer S., O’Connor T. 2014. *HTML5. W3C*.
- 574 Houghton J., Rasmussen B., Sheehan P., Oppenheim C., Morris A., Creaser C., Greenwood H., Summers
575 M., Gourlay A. 2009. Economic implications of alternative scholarly publishing models: Exploring the

- 576 costs and benefits.
- 577 International Organization for Standardization. 2013. ISO 32000-1:2008 - Document management –
578 Portable document format – Part 1: PDF 1.7. *ISO*.
- 579 International Organization for Standardization. 2014. ISO/IEC 10646:2014 - Information technology –
580 Universal Coded Character Set (UCS). *ISO*.
- 581 Kielhorn A. 2011. Multi-target publishing—Generating ePub, PDF, and more, from Markdown using
582 pandoc. *TUGboat-TeX Users Group* 32:272.
- 583 Kluyver T., Ragan-Kelley B., Pérez F., Granger B., Bussonnier M., Frederic J., Kelley K., Hamrick J.,
584 Grout J., Corlay S., others. 2016. Jupyter notebooks—a publishing format for reproducible computational
585 workflows. In: *Positioning and power in academic publishing: Players, agents and agendas*. 87–90.
586 DOI: 10.3233/978-1-61499-649-1-87.
- 587 Lamport L. 1994. *LaTeX: A Document Preparation System*. Reading, Mass: Addison-Wesley Profes-
588 sional.
- 589 Lanthaler M., Gütl C. 2012. On using JSON-LD to create evolvable RESTful services. In: *Proceedings*
590 *of the third international workshop on RESTful design*. ACM, 25–32.
- 591 Leonard S. 2016. *Guidance on Markdown: Design Philosophies, Stability Strategies, and Select Regis-*
592 *trations*. RFC Editor; Internet Request for Comments.
- 593 National Information Standards Organization. 2012. *JATS: Journal Article Tag Suite*.
- 594 Ngo T. 2006. *OFFICE OPEN XML OVERVIEW ECMA TC45*. Ecma International.
- 595 Ovidia S. 2014. Markdown for Librarians and Academics. *Behavioral & Social Sciences Librarian*
596 33:120–124. DOI: 10.1080/01639269.2014.904696.
- 597 R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna, Austria: R
598 Foundation for Statistical Computing.
- 599 Raggett D., Hors AL., Jacobs I., Le Hors A., Raggett D., Jacobs I. 1999. *HTML 4.01 Specification*. W3C.
- 600 Shotton D. 2010. CiTO, the Citation Typing Ontology. *Journal of Biomedical Semantics* 1:S6. DOI:
601 10.1186/2041-1480-1-S1-S6.
- 602 Simonsen K. 1992. *Character Mnemonics & Character Sets*. Rationel Almen Planlaegning; Internet
603 Request for Comments.
- 604 Smith AM., Katz DS., Niemeyer KE. 2016. Software Citation Principles. *PeerJ Computer Science* 2:e86.
605 DOI: 10.7717/peerj-cs.86.
- 606 Solomon D., Björk B-C. 2016. Article Processing Charges for Open Access Publicationthe Situation for
607 Research Intensive Universities in the USA and Canada. *PeerJ* 4:e2264. DOI: 10.7717/peerj.2264.
- 608 Suber P. 2012. *Open Access*. Cambridge, Mass: The MIT Press.
- 609 Towbin H., Staehelin T., Gordon J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to
610 nitrocellulose sheets: Procedure and some applications. *Proceedings of the National Academy of Sciences*
611 76:4350–4354.
- 612 Van Noorden R. 2012. Journal Offers Flat Fee for “all You Can Publish”. *Nature News* 486:166. DOI:
613 10.1038/486166a.
- 614 Van Noorden R. 2013. Open Access: The True Cost of Science Publishing. *Nature* 495:426–429. DOI:
615 10.1038/495426a.
- 616 Van Noorden R. 2014. The arXiv Preprint Server Hits 1 Million Articles. *Nature News*. DOI: 10.1038/na-
617 ture.2014.16643.
- 618 Volmer DA., Stokes CS. 2016. How to Prepare a Manuscript Fit-for-Purpose for Submission and Avoid

- 619 Getting a “desk-Reject”. *Rapid Commun. Mass Spectrom.*:n/a–n/a. DOI: 10.1002/rcm.7746.
- 620 Willinsky J. 2005. The Unacknowledged Convergence of Open Source, Open Access, and Open Science.
621 *First Monday* 10. DOI: 10.5210/fm.v10i8.1265.
- 622 Woelfle M., Olliaro P., Todd MH. 2011. Open Science Is a Research Accelerator. *Nat Chem* 3:745–748.
623 DOI: 10.1038/nchem.1149.
- 624 Yergeau F. 2003. *UTF-8, a transformation format of ISO 10646*. Alis Technologies.
- 625 Youngen GK. 1998. Citation Patterns to Traditional and Electronic Preprints in the Published Literature.
626 *Coll. res. libr.* 59:448–456. DOI: 10.5860/crl.59.5.448.