

Introducing the Scholarly Markdown Bundle

Martin Fenner, Gobbledygook

April 23, 2015

Using Markdown to author scholarly documents is an attractive alternative to the standard authoring tools Microsoft Word and LaTeX. The feeling shared by many is that Scholarly Markdown is 80% there, and that more effort is needed for the remaining 20% - moving markdown from a niche into the mainstream. What is mainly needed is building tools that connect the existing tools and ideas, resulting in one or more services attractive to a critical number of users. But maybe we also need to rethink the essential parts of Scholarly Markdown. In this post I propose that we expand the concept and define the *Scholarly Markdown Bundle*.

It is becoming increasingly clear that scholarly work can't be adequately described in a single text document, most commonly the journal article. Not only are there associated metadata, assets such as figures and supplementary information, but also the research data and software needed to produce the work described in the publication. The obvious next step is to think of scholarly work as a collection of objects, most clearly described by Carol Goble and others as Research Object Bundle.

There will probably never be a single authoring tool and format that pleases everyone. Markdown has particular inherent strengths and weaknesses, complex math or tables will probably always be easier with other formats. The strength of markdown is the simplicity of the format. Some things are hard or impossible to do, but many other things are much simpler. Creating a useful markdown editor is much easier than a word processor reading/writing `docx` format. Markdown is also a perfect format to work with version control systems such as git.

This low barrier of entry makes markdown perfect to be integrated into many workflows. And we can go one step further than ePub and Research Object Bundle, which use the related Universal Container Format (UCF) and ePub Open Container Format (OCF), respectively. Instead of using zip to compress a folder into a single file we can use git version control instead: git provides the commands `git bundle` and `git archive` to compress a project under version control with or without version history. I feel this format is both more powerful. So I propose the *Scholarly Markdown Bundle*:

- a git repository with one or more markdown files, either as a folder, or

compressed into a single file using `git bundle`

- a particular flavor or markdown called Scholarly Markdown, and discussed here and elsewhere before
- a `citeproc.json` file in the root of the project that contains all metadata relevant to the container, including references

The `citeproc.json` file is similar to the minimal metadata schema `codemeta` proposed by Matt Jones and others, but is in the format used by Pandoc today. This is important because it adds citation parsing support out of the box. The last two points rely on the Pandoc document conversion tool, so Scholarly Markdown bundles are really **markdown + Pandoc + Citeproc/CSL + git**. The format is flexible enough to not only describe scholarly articles, but also other kinds of scholarly works, including scientific software managed with git version control. And it integrates nicely with a number of existing workflows, e.g. an R project using RStudio for both code and text (in Rmarkdown). This format should also work for blogs like this one, but I would have to separate the blog posts from the Jekyll site generator code, a direction I suggested in the last post.