

Template to prepare preprints and manuscripts using markdown and github actions

Michael D. Catchen^{1,2}

¹ McGill University ² Québec Centre for Biodiversity Sciences

Correspondance to:

Michael D. Catchen — michael.catchen@mail.mcgill.ca

Purpose: This template provides a series of scripts to render a markdown document into an interactive website and a series of PDFs.

Motivation: It makes collaborating on text with GitHub easier, and means that we never need to think about the output.

Internals: GitHub actions and a series of python scriipts. The markdown is handled with pandoc.

- 1 This template uses pandoc (and a few additional python glue scripts) to facilitate the production of
 - 2 scientific articles using a standard markdown file. The objective is to ensure that standard markdown
 - 3 (with the important exception of the pandoc-crossref citation markup) will be rendered into an
 - 4 interactive website (which allows collaborative annotations with the `hypothes.is` platform), a “draft”
 - 5 style PDF (double-spaced, numbered lines, figures at the end), and a “preprint” style PDF (with slightly
 - 6 more reader-friendly pagination).
- 7 The core bit of configuration is the `metadata.json` file, which handles information about authorship,
 - 8 affiliations, the abstract, keywords, etc. All documents will be deployed to `gh-pages` *only* on push events
 - 9 from the `main` branch. All of the artifacts will be built when doing pull requests, so you can check that
 - 10 merging a branch is *not* going to cause the compilation of the documents to fail; indeed, you can download
 - 11 the artifacts produced during the run, to check the PDF and html files. The website is only updated from
 - 12 the `main` branch.
- 13 The workflow is *very* GitHub based, and so the manuscript file is the `README.md` - this is not going to be a
 - 14 huge issue as 90% of the markdown is standard, with the exception of the citations and mathematics, so
 - 15 this will render (mostly) like a normal `README` file.

16 Deploying the template

- 17 The process of deploying this template has been *greatly* streamlined from previous versions:
 - 18 • Click on the “Use this template” button
 - 19 • Edit `README.md` with your own text, commit, and push
 - 20 • This push will trigger the first build - the builds are only active on the `main` branch (*not master!*),
 - 21 and on pull requests
 - 22 • Go to `http://you.github.io/repo-name/` to view the html version, and get access to the PDFs
 - 23 • Add your references to the `references.bib` file
 - 24 • Edit the `metadata.json` file to add the title, abstract, authors
- 25 In particular, note that *you do not need* to create a personnal access token to deploy to `gh-pages` (from
 - 26 where the website is served).

27 **The metadata file**

28 **General information**

29 The title is a field in the `metadata.json`:

```
{  
    "title": "Preprint template"  
}
```

30 **Authorship**

31 Authors are listed as objects in the `authors` block. Each author is specified as follows:

```
{  
    "familyname": "Bob",  
    "givennames": "Alice",  
    "email": "alice.bob@u.edu",  
    "orcid": "0000-0000-0000-0001",  
    "affiliations": [  
        "Affiliation 1",  
        "Affiliation 2"  
    ],  
    "status": ["corresponding", "equal"]  
}
```

32 The `email` field is recommended for all authors. The `status` field is only useful for the corresponding

33 author, and to denote equal contributions. These informations are rendered on the initial page. If an

34 `orcid` is given, it will be linked on the HTML and PDF versions.

35 Note that there is *no need* to number the affiliations - a small python script will take care of this

36 automatically.

37 **Abstract**

- 38 This template supports three types of abstracts, indicated in the metadata file as `abstract`:
- 39 A regular abstract is defined as

```
"abstract": "A very long string"
```

- 40 An itemized abstract is an array of strings, each representing a bullet point:

```
"abstract": [
    "Point 1",
    "Point 2"
]
```

- 41 A structured abstract is an object with key-value pairs :

```
"abstract": {
    "Location": "Worldwide",
    "Organisms": "Mammals"
}
```

42 **Citation style**

- 43 The `citationstyle` key corresponds to the name, with `.csl` omitted, of a CSL stylesheet stored in the [citation style language](#) repository. Note that there is no difference between main and dependent styles, the 44 build engine will take the correct steps to get the correct style. The default is `"citationstyle":`
- 45 `"ecology-letters"`. There is a longer section about references management later on.

47 **References management**

- 48 The references are managed by pandoc. Note that we *do not* use `pandoc-citeproc`, which was an external 49 module for older pandoc versions. References *must* be stored in a `references.bib` file, and that it would 50 make sense to order it alphabetically by key.

51 We use [Zotero](#) for references management, and for the lab's manuscripts, we work from folders in a shared
52 library (with a folder for every manuscript).

53 It is recommended to use the [Better BibTeX](#) plugin for citation key generation, and auto-export of the
54 shared library to the `references.bib` file. We use a citation key format meant to convey information on
55 the author (first author full name), date (complete year), and title (first three letters of the first two non-stop
56 words). It must be set in the Better BibTeX preferences as (you might need to remove the line changes):

```
57 [auth:fold]  
58 [year]  
59 [title:fold:nopunctordash:skipwords:lower:select=1,1:substring=1,3:capitalize]  
60 [title:fold:nopunctordash:skipwords:lower:select=2,2:substring=1,3:capitalize]
```

61 It is a good idea to configure Better BibTeX to auto-export on change, and to remove a lot of fields that are
62 not strictly speaking required for references. The list of fields we usually ignore is:

63 `abstract, copyright, annotation, file, pmid, month, shorttitle, keywords`

64 The citations are done using the normal markdown syntax, where `@Elton1927AniEco` produces Elton
65 (1927), and `[@Camerano1880EquViv]` produces (Camerano 1880).

66 **Figures, Tables, and other floats**

67 Note that you can wrap the text of legends for both figures and tables. This avoids the issue of having very
68 long lines.

69 **Mathematics**

70 The following equation

$$J'(p) = \frac{1}{\log(S)} \times \left(- \sum p \times \log(p) \right) \quad (1)$$

71 is produced using

```
 $$J'(p) = \frac{1}{\text{log}(S)} \times \dots $$ {#eq: eq1}
```

⁷² and can be referenced using @eq: eq1, which will result in eq. 1. Note that because we use
⁷³ pandoc-crossref, the label “eq.” will be generated automatically.

⁷⁴ Tables

⁷⁵ Table legends go on the line after the table itself. To generate a reference to the table, use {#tbl: id} –
⁷⁶ then, in the text, you can use {@tbl: id} to refer to the table. For example, the table below is tbl. 1. You
⁷⁷ can remove the *table* in front by using !@tbl: id, or force it to be capitalized with *tbl: id.

Table 1: This is a table, and its identifier is id – we can refer to it using {@tbl: id}. Note that even if the table legend is written below the table itself, it will appear on top in the PDF document.

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

⁷⁸ Figures

⁷⁹ Figures can have a legend – all figures *must* be in the `figures/` folder of the project, as it is also used for
⁸⁰ the website. We recommend to use good resolution images, rather than PDFs, or at least to have multiple
⁸¹ versions available.

⁸² ! [This is the legend of the figure...] (figures/figure.png){#fig: figure}

⁸³ [Figure 1 about here.]

⁸⁴ We can now use @fig: figure to refer to fig. 1.

85 **Example text**

86 Connectance, defined as the ratio of realized interactions on the total number of potential interactions, is
87 one of the most common descriptor of network structure. In a bipartite network with T species at the top,
88 and B at the bottom, having a total of L interactions, it is defined as $Co = L/(T \times B)$. Connectance has a
89 lower bound, as the network cannot have fewer interactions than the number of species in its more
90 speciose level – the minimal connectance is therefore $c_m = \max(T, B)$. This makes the connectance of
91 networks of different sizes difficult to compare, especially since bipartite networks tends to have a low
92 connectance. For this reason, we used a corrected version of connectance, defined as

$$Co^* = \frac{L - c_m}{T \times B - c_m}. \quad (2)$$

93 **This is a subsection**

94 This takes values between 0 (the network has the minimal number of interactions) and 1 (all species are
95 connected), but is robust to variations in species richness.

96 **This is another subsection**

97 This takes values between 0 (the network has the minimal number of interactions) and 1 (all species are
98 connected), but is robust to variations in species richness.

99 **Some non-standard maths**

100 The phylogenetic reconstruction of $\hat{\mathcal{L}}$ and $\hat{\mathcal{R}}$ has an associated uncertainty, represented by the breadth of
101 the uniform distribution associated to each of their entries. Therefore, we can use this information to
102 assemble a *probabilistic* metaweb in the sense of (**Poisot2016StrPro?**), *i.e.* in which every interaction is
103 represented as a single, independent, Bernoulli event of probability p .

104 Specifically, we have adopted the following approach. For every entry in $\hat{\mathcal{L}}$ and $\hat{\mathcal{R}}$, we draw a value from
105 its distribution. This results in one instance of the possible left ($\hat{\ell}$) and right (\hat{r}) subspaces for the
106 Canadian metaweb. These can be multiplied, to produce one matrix of real values. Because the entries in

¹⁰⁷ $\hat{\ell}$ and \hat{r} are in the same space where \mathcal{L} and \mathcal{R} were originally predicted, it follows that the threshold ρ
¹⁰⁸ estimated for the European metaweb also applies. We use this information to produce one random
¹⁰⁹ Canadian metaweb, $N = \hat{\mathcal{L}}\hat{\mathcal{R}}' \geq \rho$.

¹¹⁰ Because the intervals around some trait values can be broad (in fact, probably broader than what they
¹¹¹ would actually be, see e.g. **Garland1999IntPhy?**), we repeat the above process 2×10^5 times, which
¹¹² results in a probabilistic metaweb P , where the probability of an interaction (here conveying our degree of
¹¹³ trust that it exists given the inferred trait distributions) is given by the number of times where it appears
¹¹⁴ across all random draws N , divided by the number of samples. An interaction with $P_{i,j} = 1$ means that
¹¹⁵ these two species were predicted to interact in all 2×10^5 random draws, etc..

¹¹⁶ References

- ¹¹⁷ Camerano, L. (1880). Dell'equilibrio dei viventi merce la reciproca distruzione. *Atti Della R. Accad. Delle*
¹¹⁸ *Sci. Torino*, 15, 393–414.
- ¹¹⁹ Elton, C.S. (1927). *Animal ecology*. University of Chicago Press.

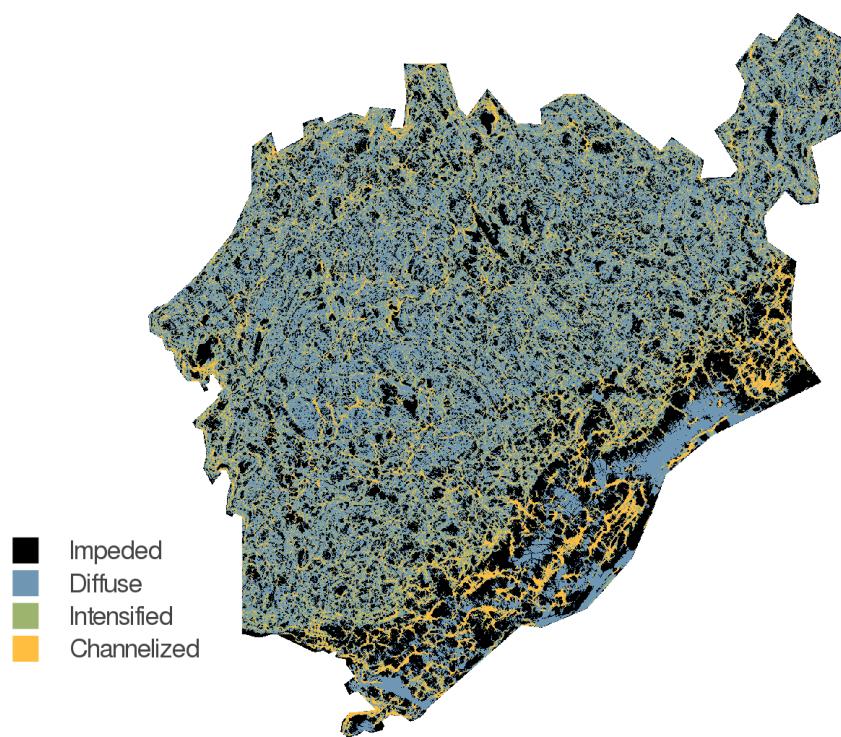


Figure 1: This is the legend of the figure, which will be shown in the margin in preprint mode, and underneath the figure in draft mode. The legend can contain references, etc. It is advised to use a resolution of at least 600dpi for the figures.