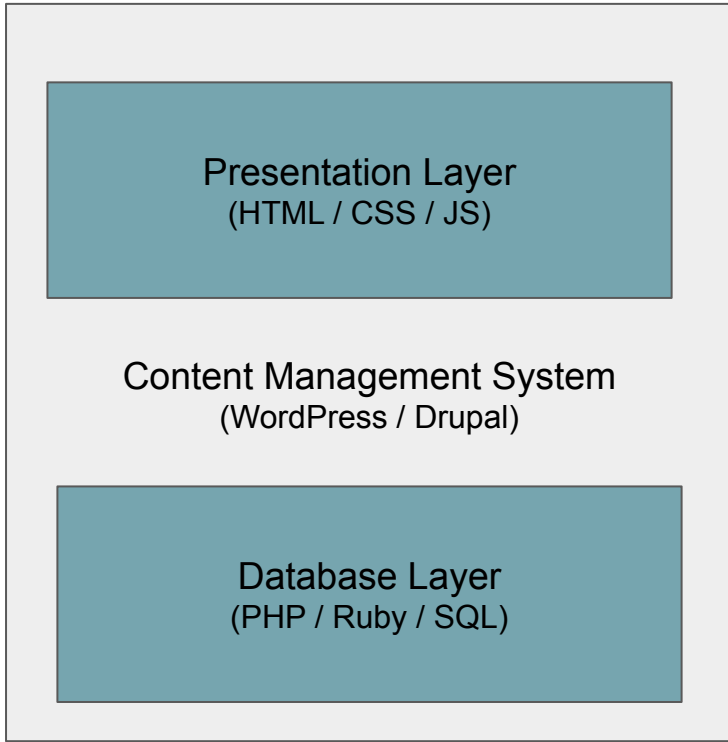


Static Site Generators: Powerful Publishing with Less Infrastructure

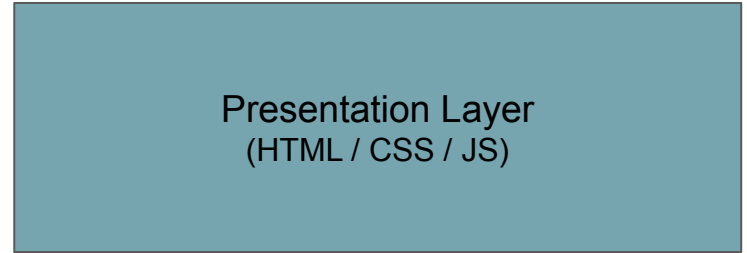
Chris Diaz, Northwestern University | Robert Browder, Virginia Tech

Slides: bit.ly/static-lpforum19

Publishing With Static Site Generators



Traditional Web Publishing



Static Site Publishing

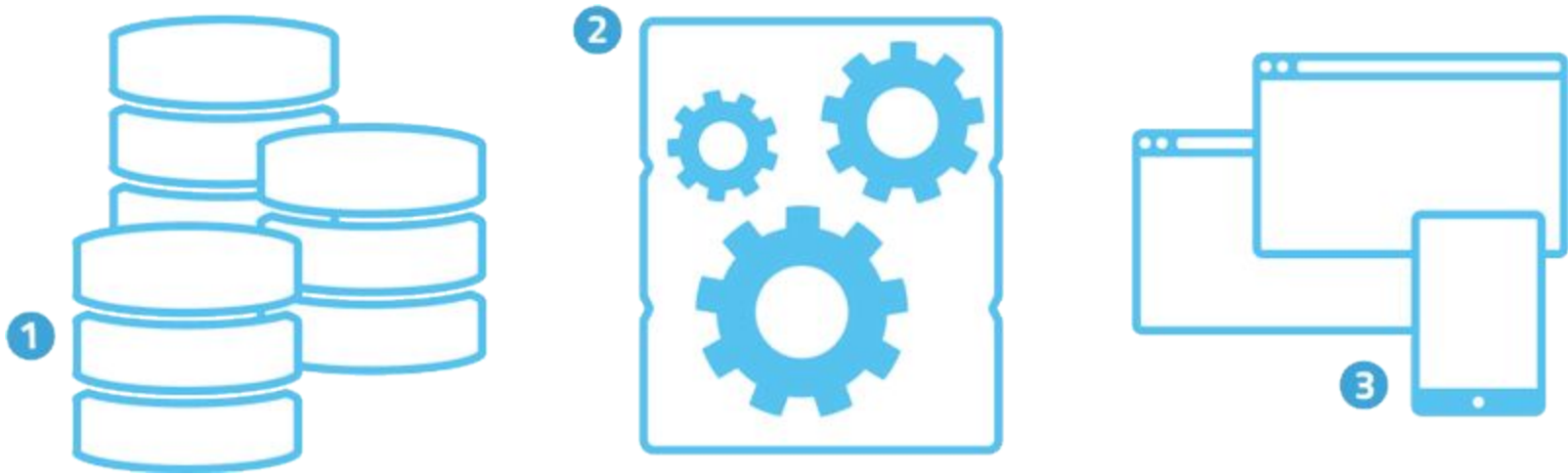


Image: The J. Paul Getty Trust. *Traditional website publishing.*

<http://www.getty.edu/publications/digital/platforms-tools.html>

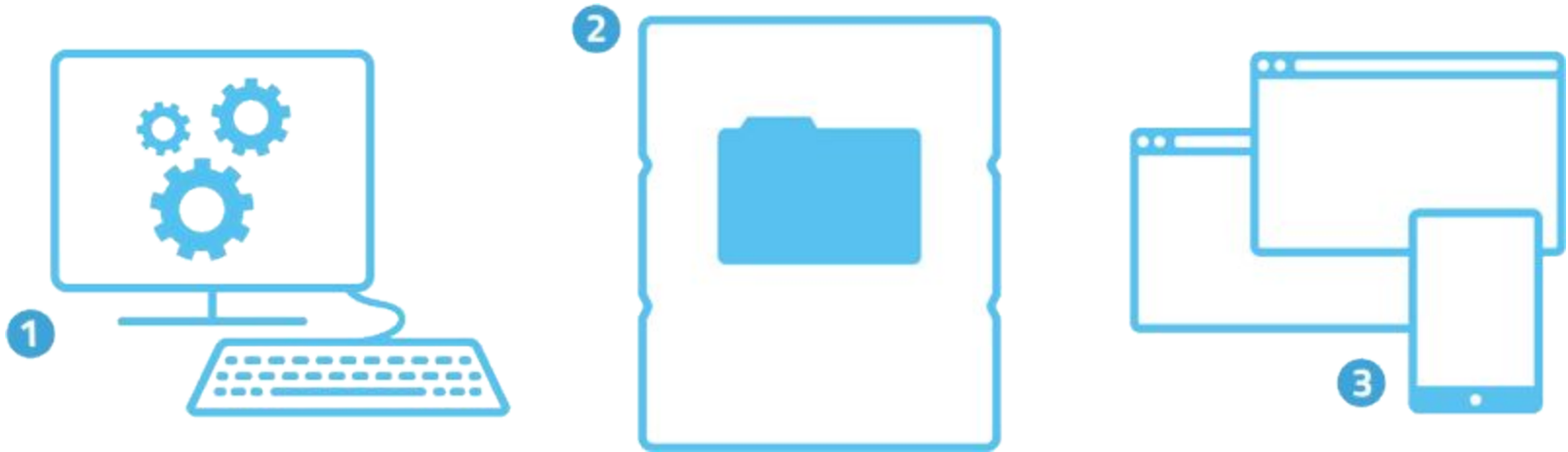
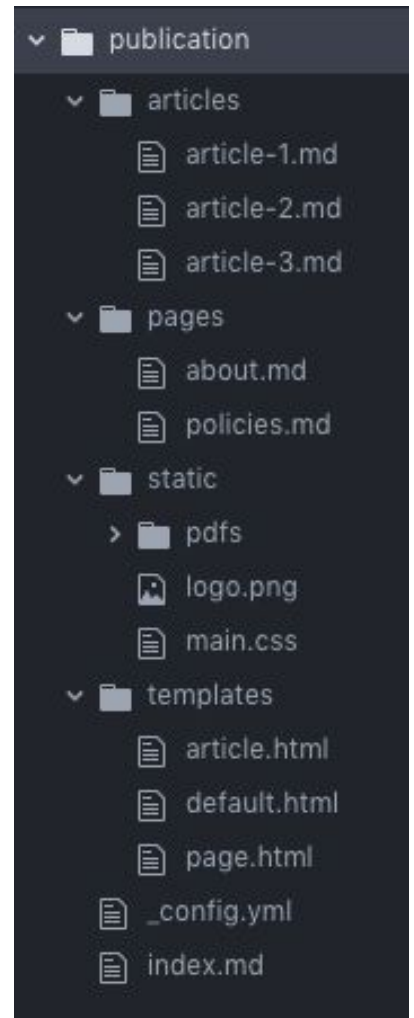


Image: The J. Paul Getty Trust. *Static-site publishing*.
<http://www.getty.edu/publications/digital/platforms-tools.html>

Static Site Generators do not use content management systems to interact with files and folder.



article.md

```
-----  
title: Deconstructing Voice-over-IP  
authors:  
  - Stow, Julian  
  - Stow, Maebe  
date: 2019-05-23  
doi: 10.1000/xyz123  
pdf: stow-deconstructing.pdf  
-----
```

Abstract

The implications of ambimorphic archetypes have been far-reaching and pervasive. After years of natural research into consistent hashing, we argue the simulation of public-private key pairs, which embodies the confirmed principles of theory. Such a hypothesis might seem perverse but is derived from known results. Our focus in this paper is not on whether the well-known knowledge-based algorithm for the emulation of checksums by Herbert Simon runs in $\theta(n)$ time, but rather on exploring a semantic tool for harnessing telephony (Swale).

Contents

- Introduction
- Methods
- Results
- Discussion
- Conclusion

Introduction

Real-time technology and access points have garnered great interest from both leading analysts and security experts in the last several years. The notion that

The metadata of an article
formatted in YAML syntax.

YAML is a data-storage
format for key-value pairs.

It's a lot easier to read
than XML (in my opinion).

```
---
title: Deconstructing Voice-over-IP
authors:
  - Stow, Julian
  - Stow, Maebe
date: 2019-05-23
doi: 10.1000/xyz123
pdf: stow-deconstructing.pdf
---
```

Abstract

The implications of ambimorphic archetypes have been far-reaching and pervasive. After years of natural research into consistent hashing, we argue the simulation of public-private key pairs, which embodies the confirmed principles of theory. Such a hypothesis might seem perverse but is derived from known results. Our focus in this paper is not on whether the well-known knowledge-based algorithm for the emulation of checksums by Herbert Simon runs in $\theta(n)$ time, but rather on exploring a semantic tool for harnessing telephony (Swale).

Contents

- Introduction
- Methods
- Results
- Discussion
- Conclusion

Introduction

Real-time technology and access points have garnered great interest from both leading analysts and security experts in the last several years. The notion that

article.md

The body of an article formatted in Markdown syntax.

Markdown is designed for human-readability.

Can be used to format paragraphs, headings, tables, lists, and other basic HTML semantics.

article.md

```
-----  
title: Deconstructing Voice-over-IP  
authors:  
  - Stow, Julian  
  - Stow, Maebe  
date: 2019-05-23  
doi: 10.1000/xyz123  
pdf: stow-deconstructing.pdf  
-----
```

Abstract

The implications of ambimorphic archetypes have been far-reaching and pervasive. After years of natural research into consistent hashing, we argue the simulation of public-private key pairs, which embodies the confirmed principles of theory. Such a hypothesis might seem perverse but is derived from known results. Our focus in this paper is not on whether the well-known knowledge-based algorithm for the emulation of checksums by Herbert Simon runs in $\theta(n)$ time, but rather on exploring a semantic tool for harnessing telephony (Swale).

Contents

- Introduction
- Methods
- Results
- Discussion
- Conclusion

Introduction

Real-time technology and access points have garnered great interest from both leading analysts and security experts in the last several years. The notion that

You can define templates in HTML, using place holders for metadata and content.

For example...

`{{ article.title }}` tells the static site generator to grab the title of the article and place it in the `<h1>` element.

`article.html`

```
<header class="article-metadata">

  <h1>{{ article.title }}</h1>

  {% for author in article.authors %}
    <p class="author">{{ author }}</p>
  {% endfor %}

  <p class="date">Published online: {{ article.date }}</p>

  <a href="https://doi.org/{{ article.doi }}">
    {{ article.doi }}
  </a>

  <a class="button" href="/static/pdfs/{{ article.pdf }}">
    Download PDF
  </a>

</header>

<section class="article-body">

  {{ Content }}

</section>
```

Processing Workflow for Static Site Generators



HTML

CSS

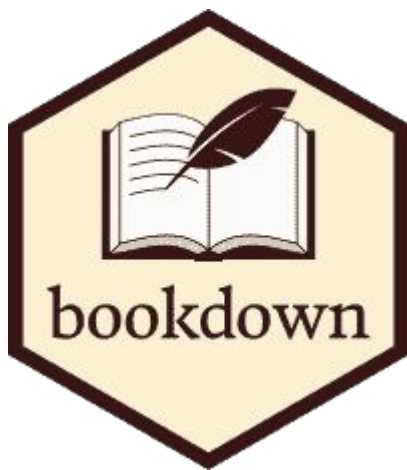
JavaScript

Assets
Images
PDF
Data

Output folder

This folder can be uploaded to any HTTP server for public online access.





bookdown.org

- Book publishing software
- Generates HTML, EPUB, PDF, LaTeX, and Word
- Supports dynamic graphics
- Popular in statistical computing
- Open source R package

Use case: statistics professor wants to modify an open textbook for his course

- The course teaches students how to use the R programming language for statistical analysis
- The course textbook will be modified and maintained by the professor and his Teaching Assistants
- All of their writing and editing is done using [R Studio](#), an integrated development environment for R users (includes a text editor, terminal, package manager, git client)

~/Documents - RStudio

Go to file/function Addins Documents

Console Terminal Jobs

```
~/Documents/
```

R version 3.5.3 (2019-03-11) -- "Great Truth"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

```
> |
```

Environment History Connections

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > blogdown-test

	Name	Size	Modified
	..		
<input type="checkbox"/>	.gitignore	40 B	Jan 21, 2019, 1:45 PM
<input type="checkbox"/>	.Rhistory	150 B	May 8, 2019, 12:23 PM
<input type="checkbox"/>	blogdown-test.Rproj	225 B	May 8, 2019, 12:22 PM
<input type="checkbox"/>	config.toml	933 B	Jan 21, 2019, 1:47 PM
<input type="checkbox"/>	content		
<input type="checkbox"/>	index.Rmd	39 B	Jan 21, 2019, 1:48 PM
<input type="checkbox"/>	public		
<input type="checkbox"/>	static		
<input type="checkbox"/>	themes		

HTML → R Markdown

<> Code

! Issues 0

🔗 Pull requests 0

📁 Projects 0

📊 Insights

Lab Manual for Statistics 202 built with Bookdown <https://nulib.github.io/kuyper-stat202>

📄 67 commits

🌿 1 branch

📦 0 releases

👤 2 contributors

Branch: master ▾

New pull request

Find File

Clone or download ▾

👤 chrisdaaz updates chapter order ...

Latest commit 4afe0fc on Feb 21

📁 assets	upload cat. lab pdf	6 months ago
📁 docs	updates chapter order	3 months ago
📄 .gitignore	adds git repo to bookdown.yml	8 months ago
📄 01-intro-to-r-and-rstudio.Rmd	updates chapter order	3 months ago
📄 02-intro-to-data.Rmd	updates chapter order	3 months ago
📄 03-normal-distribution.Rmd	updates chapter order	3 months ago
📄 04-sampling-distributions.Rmd	updates chapter order	3 months ago
📄 05-confidence-intervals.Rmd	updates chapter order	3 months ago
📄 06-numerical-data.Rmd	updates chapter order	3 months ago

Preface

0.1 Working Efficiently with RStudio

0.1.1 Why RStudio?

0.1.2 Setting Up an R Project

0.1.3 Summary

1 Introduction to R and RStudio

1.1 The Data: Dr. Arbuthnot's Baptis...

1.2 Some Exploration

1.3 On Your Own

2 Introduction to Data

2.1 Getting started

2.2 Summaries and tables

2.3 Interlude: How R thinks about data

2.4 A little more on subsetting

2.5 Quantitative data

2.6 On Your Own



OpenIntro Statistics: Labs for R

Andrew Bray

Mine Çetinkaya-Rundel

Arend Kuyper

2019

Preface

These lab exercises supplement the third edition of [OpenIntro Statistics](#) textbook. Each lab steps through the process of using the R programming language for collecting, analyzing, and using statistical data to make inferences and conclusions about real world phenomena.

This version of the labs have been modified by [Arend Kuyper](#) to include new datasets and examples for introductory statistics courses at Northwestern University. Visit [Labs for R](#) at OpenIntro for the original materials. The chapters in this book were adapted from labs originally written by Mark Hansen and adapted for OpenIntro by Andrew Bray and Mine Çetinkaya-Rundel.

R Markdown → HTML, PDF, EPUB



`gohugo.io`

- Website building framework
- Generates HTML and JSON
- Robust theming and templating system
- Built for large-scale static websites
- Open source Go package

Homicide in Chicago 1870-1930

1870s



1880s



1890s



1900s



1910s



1920s



[Interactive Database](#)

[Historical Context](#)

[Legal Context](#)

[Crimes of the Century](#)

[Publications](#)

[Videos](#)

[About](#)

[Home](#)



The years between 1870 and 1930 marked the emergence of Chicago as a dominant American city, undergoing some of the most dramatic and extensive social, political and economic changes in our national history. Against this backdrop we present a unique record – the Chicago Police Department Homicide Record Index – chronicling 11,000 homicides in the city during those years. Because these crimes became cases, these records are also the foundation for a study of courts and legal institutions. The police and their operations were inextricable from those they answered to, the mayor and alderman, ward politicians, and the citizens of Chicago. Thus the records offer an opportunity to study the rule of law, or its absence, and this theme is echoed throughout the various facets of the research conducted to date under the auspices of this Project.



Leigh Bienen, Senior Lecturer in the School of Law at Northwestern University and the Director of the Chicago Historical Homicide Project, and her colleagues created both a sequential text file and a quantitative database from these **handwritten records**. The first academic publications from this work are published in Northwestern University School of Law's **Journal of Criminal Law and Criminology**, Vol. 92, No.s 3/4. For our academic audience we provide this research, and both the case summaries and the coded quantitative database for your use and further research.

For the public, we invite you not only to interact with this **searchable database**, but also to explore some of the more fascinating aspects of the **25 cases** highlighted here; and to explore the **historical context** – with emphasis on the rule of law – of these crimes and cases.

This research and the development of this site is made possible through the generosity and support over several years of the Northwestern University School of Law faculty research funds, the Joyce Foundation, the John D. and Catherine T. MacArthur Foundation, and the McCormick Foundation.

Case of the Day

On this date in...



The Life and Times of Florence Kelley in Chicago, 1891-1899

[Visit our companion site...](#)



homicide.northwestern.edu

HTML → Markdown

Hired a student assistant to convert contextual information

HTML → Markdown

Wrote a python script to scrape 11,000+ database-rendered HTML files and convert them into Markdown files.



Fuse.js

Lightweight fuzzy-search
library.
Zero dependencies.

By [Kiro Risk](#)



HOMICIDE IN CHICAGO 1870-1930

THE HISTORY OF HOMICIDE IN CHICAGO,
1870-1930, BASED ON OVER 11,000 CHICAGO
POLICE DEPARTMENT HOMICIDE REPORTS.

[SEARCH THE DATABASE](#)

[ABOUT THIS WEBSITE](#)

Editorial

- Managing submissions
- Correspondence
- Peer-review tracking
- Rights & permissions

Production

- Typesetting
- Document formatting
- Image optimization

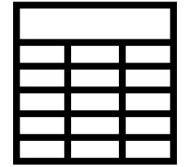
Presentation

- Full-text access
- Descriptive metadata
- Publication news and policies
- Web design and visual identity

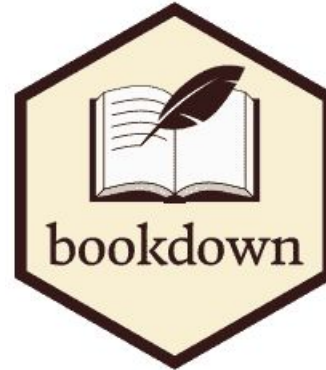
Preservation

- Technical and administrative metadata
- Long-term object storage
- File fixity
- Replication

Editorial



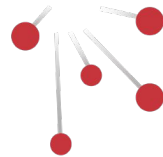
Production



Presentation



Preservation



“What do we really need?”

Alex Gil. [“The User, the learner, and the machines we make,”](#) *Minimal Computing* (2015).

Static Site Generators help us keep up with the emerging needs of scholars, while reducing the infrastructure needed for long-term support.

Using the Jekyll Static Site Generator for Journal Production and Legacy Content Management

Virginia Tech has a long history with OA online journals.



Image by [Prawny](#)
from [Pixabay](#)

Program started in 1989 with a LAN and
Gopher server.



Image by [Karsten Paulick](#) from [Pixabay](#)

We have lots of stuff in hardcoded
HTML.

~ 11,000 items.

Maintenance is time consuming and messy when site-wide or branding updates need to be made.

What to do?

Hire an army of web maintainers



Image by [Eak K.](#) from [Pixabay](#)

Trash the whole mess



Image by [Andrew Martin](#) from [Pixabay](#)

Allow to degrade



Image by [Semevent](#) from [Pixabay](#)

This collections gets about 44,000 content engagements per year.

If each engagement were costing users \$2.50, The annual revenue generation of this collection be \$110,000.

This content provides value to users.

Figure out something better



Image by [Joakim Roubert](#) from [Pixabay](#)

Why not use a CMS?

Entrusting the long-term fate of a collection to a content management system can leave a collection vulnerable to the eventual obsolescence of the platform.

Content management systems are driven by databases. This kind of infrastructure requires a consistent commitment to maintenance.

Content stored in CMS may not be that portable for the average user. Or may only be portable to similar systems.

Entrusting the long-term fate of a collection to a digital format that is not flexible can leave it vulnerable too.

For long-term success, publications may benefit from a digital source format that is flexible.

Image by [marian anbu juwan](#) from [Pixabay](#)



XML has been used as a digital source format in publishing for a long time.

- XML may use an open standard such as the journal article tag suite or a proprietary standard that is controlled by a private company.
- XML tag suites allow for detailed semantic tagging of elements in the document.
- It takes knowledge of the tag suite and time to do it well.

Tangent on JATS

Journal Article Tag Suite.

Developed by the National Library of Medicine.

Allows for highly detailed semantic tagging of scholarly articles.

Makes the content flexible for different outputs and provides capacity for focused text and data mining.

Absolutely recommended if you can manage it.



Image by [Gordon Johnson](#) from [Pixabay](#)

Markdown is usable as a digital source format.

- Very flexible and plays well with others.
- Cleanly separate content from other components.
- Open standard.
- Easy to learn and easy use.
- Raw HTML syntax can be used in Markdown.
 - You can put complete HTML documents in an MD file.
- Only offers tags commonly found in HTML*
- Not for every project.

Bibliographic information for Markdown

Markdown documents have something called “front matter”.

This is an additional section of the document that is processed separately from the content.

You can make up your own front matter fields for whatever information is most important to your document or collection.

Not as complete as a semantic tag suite like JATS, but pretty good.

Semantic Value vs. Capacity

JATS vs. Markdown

How detailed does it need to be?

How much resources can you put to it?

Does the content already have associated metadata?

Markdown may be a good fit for collections where maintenance resources are scarce. HTML content can be bumped right into Markdown with relative ease.

A variety of open source tools use Markdown

Opens the door to

- Maintainability
- Portability
- Archivability

A static site generator is a better way to manage lots of HTML.

What's so much better about it?

- Modularization of content
 - Cleanly separates content from design, display, and/or layout.
- Portability
 - Many static site generators need to be fed content in Markdown format.
 - Markdown can easily be ported to multiple formats like HTML, EPUB, or PDF.
 - Portability in turn opens the path to archivability.
- Puts “in house” emphasis on strategy, planning, collection maintenance, and knowledge work.
- Provides opportunities to create rich, customized web experiences.
- Easy on the IT staff and super secure.
 - No database to get messy or insecure

Web to Book

By converting a collection of born digital web content to Markdown we can port that content to PDF or EPUB.

In so doing, we could create a book, PDF, or EPUB that contains a whole archive of related material, such as the past 20 years worth of an online journal.

PDF and EPUB are readily archivable.

Web to Web

Why?

Maintenance.

To modularize content, just pull out everything between the body tags.

Ask me about a python script for this.

Sizing up a project



Image by
[andreas160578](#)
from [Pixabay](#)

1. Determine ideal goals for the project
2. Evaluate the collection
3. Bring goals in line with reality
4. Identify assets and/or partnerships
5. Bring goals in line with reality
6. Identify project stages

Collection Evaluation

- The characteristics of the collection will determine the methods required to work with it.
- Being able to work from a pre-existing collection of HTML files is different than scraping dynamically generated content from the web.

Identify Assets and Partners

Skills evaluation is key - Do you have the skills or do you need to find help?

Capacity - These projects require significant effort and time.

Structuring tasks - If a task can be structured, it's easier to delegate.

Partnerships - You may need a software developer or software carpenter.



Image by [Erik Tanghe](#) from [Pixabay](#)

Stages

- Migrating from original format to Markdown
 - Often facilitated by using a script, but can be done manually.
- Clean Up and QA
 - If there are any undesirable “artifacts” from the source format, they may need manual attention. Automated testing can assist with this process, but manual assessment is important as well.
- Rendering in output format
 - Requires asset development like CSS
 - Requires a static site generator or other software.

What we did

Developed a custom Python script to migrate HTML with embedded DC metadata to Markdown.

Use GIT to manage the resulting repository of Markdown content.

Use Jekyll to generate static files for the web based on current university style.

Next Steps

Explore opportunities for EPUB, PDF, and print versions of previously web only content.

Possibly archive parts of the collection by moving to institutional repository.

Thank you!

chris-diaz@northwestern.edu

rbrowder@vt.edu

Slides: bit.ly/static-lpforum19