

Publishing WG

October 2017

Tzviya Siegman
Garth Conboy
Dave Cramer
Ivan Herman

Goals

“It is the goal of the Publishing Working Group to provide, in concert with other W3C Groups ... to make the combination of traditional publishing and the Web complete in terms of the **readers’ needs**, portability, distribution, archiving, offline access, or reliable cross referencing. In short, all publications—with all their specificities and traditions—should become **first-class entities on the Web**, as Web Publications.”

<https://www.w3.org/2017/04/publ-wg-charter/#goals>

We are Document-oriented

- The main focus of traditional publishers is documents and publications, or collections of documents.
- Apps and operating systems are incidentals. We may create Learning Management Systems and platforms to distribute our products, but the focus is the documents.
- “Web of Documents” once described the Web. “Content is King” is a phrase frequently used about publishing.

State of the World Today

- EPUB (and PDF) solve many of these problems
- But... the publications are off the Web
- How does this impact the user experience? A basic hyperlink takes a user out of the publication. There often isn't even a back button. This is sometimes called “immersive reading”, but shouldn't it be the reader's choice? Should a user encountering a bibliography have the same “immersive” experience as a reader of *The Underground Railroad*?
- EPUB has some of its own issues (XML reliance, limited support, authoring, silos, cross-references)

Not everything is in EPUB

- Many people think about trade fiction and maybe text books when they hear about publishing.
- What about scholarly journals? Magazines? News? Corporate documentation? And, yes textbooks?
- In many cases, there is an HTML version, and going offline means PDF. In the world of scholarly publishing, the “document of record” is the PDF.
- Issues such as citations and archiving for scholarly publishing have prevented the adoption of EPUB.
- Ease of authoring is an issue that has probably prevented adoption of EPUB at a smaller scale

Journal Articles are HTML or PDF

 PDF
 Info
 References
 Figures



International Journal of
**QUANTUM
CHEMISTRY**

[Explore this journal >](#)

Full Paper

Periodic trends in constants of triatomic molecules

Joshua Barrow , Ken Caviness, Ray Hefferlin, Devin Nash

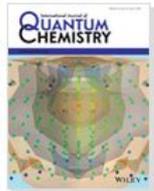
First published: 29 March 2016 [Full publication history](#)

DOI: 10.1002/qua.25137 [View/save citation](#)

Cited by (CrossRef): 0 articles [Check for updates](#) [Citation tools](#) ▼

 score 3

R. Hefferlin is deceased.



[View issue TOC](#)
Volume 116, Issue 14
July 15, 2016
Pages 1071–1083

 Text size  Share

- Abstract
- Introduction
- Recent Work in Spectroscopic Constants of Diatomic Molecules
- An Extension from Diatomic to Triatomic Molecules is Sought
- Motivation and Goals
- Methods
- Data Organization and Visualization
- Conclusions
- Acknowledgments
- References
- Related Content

Abstract

Working from the arrangements of both row and group numbers developed within Mendeleev's periodic table of elements, periodic trends can be shown to exist in many constants of triatomic molecules: an extension of the Periodic Law for atoms to the realm of molecules. Trends are identified for vibrational frequencies, bond lengths, and to a lesser extent interior angles. This work includes empirical sources for such data, supplemented with calculations using diatomic analogs

What's the problem?

- Citations refer to PDF “document of record” or one of several methods of HTML referencing. (Remember people are hired and tenured based on number of citations.)
- Navigation is a layer added to the site. It's different for each publisher's platforms. Is it accessible? How does the user access navigation?
- Offline experience? PDF only. (Try that on a small screen device...)
- Data, which is often the only part of the paper that people care about, is not part of the PDF. This is especially relevant in the Open Access world.

The Missing Link

Publications are a logical whole that may be assembled from numerous web resources. The web is organized around web pages, but there isn't a natural way to describe a collection of resources as a whole:

ch01.html

ch02.html

ch03.html

ch04.html

ch05.html

style.css

cover.png

whitewhale.jpg

book.js

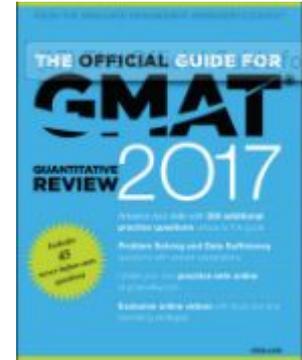
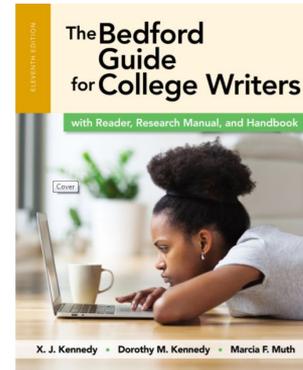
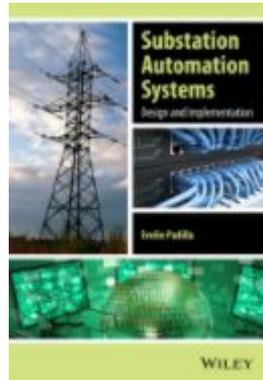
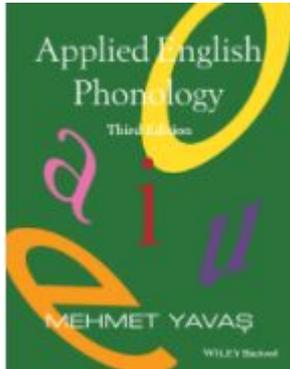
polyfill.js

Basic Requirements

- **Addressability:** one URL to access the publication, the collection of documents
- **Components:** what resources are part of the publication?
- **Sequence:** what is the default order of the primary resources?
- **Metadata:** how to describe the whole rather than the parts
- **Personalization:** how to allow the reader to adjust the publication's presentation to suit their needs
- **Annotations:** readers want to highlight, comment, bookmark
- **Offline:** the state of the network should not control the user experience

Some Sample EPUBs

- Substation Automation Systems <http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118987209.html>
- Applied English Phonology <http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118944526.html>
- The Bedford Guide for College Writers
<http://www.macmillanlearning.com/catalog/Product/bedfordguideforcollegewriterswithreaderresearchmanualandhandbook-eleventhedition-kennedy>
- The Official Guide for GMAT® 2017 <https://gmat.wiley.com/course/index/id/119>



What's the first thing you noticed?

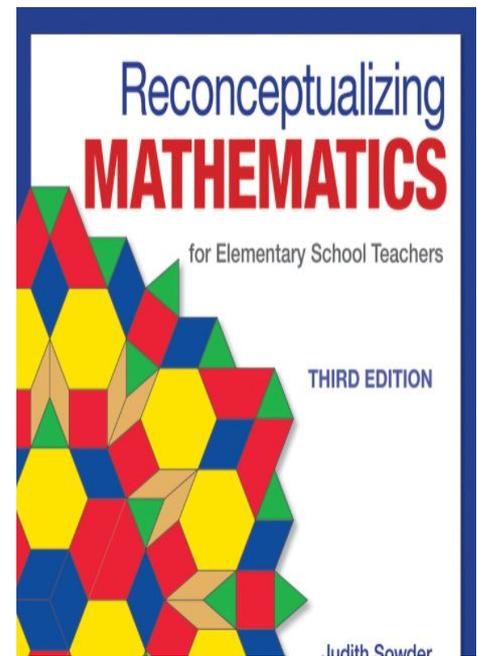
Those links go to places to buy the books!

These books do not “live” on the Web.

What happens when a user wants a publication to interact with the Web?

Web of STEM?

- MathML is a W3C standard.
- Publishers, ebook reading systems, journals platforms rely on MathJax. Without MathJax, we would be forced to revert to images. EPUB systems that do not allow for JavaScript all use images. Often the image descriptions are still provided by “” MathJax.
- It’s not enough. What’s missing?
 - Elementary math.
 - Chemistry.
 - Support.



4. Finally, use the drawing to solve the problem. The drawing shows that the total number of hot dogs eaten is $2\frac{3}{4} + 1\frac{1}{2}$, or $4\frac{1}{4}$. What part of the total was eaten by Albert?
- Albert ate $2\frac{3}{4}$ of the total ($4\frac{1}{4}$) hot dogs, which is $\frac{11}{17}$ of the hot dogs. This value is about two-thirds of the hot dogs. Is this a reasonable answer to the question?

Hyperlinks

This link in a “Further Reading” section of Substation Automation opens the user’s default browser.

How does the user get back to the book?

How does this shift focus for a user of AT?

allowed value, malfunction can be experienced.

Voltage regulation is the process that maintains the power system voltage in the permitted range. It changes the winding turn number of power transformers through a transformer subsystem called a *tap changer*. Control commands and signals associated with such a subsystem and managed by the SAS include:

- Control commands to modify tap-changer position.
- Tap-changer position indication.
- Alarms due electrical or mechanical faults on all the chain of voltage regulation process.

Further Reading

Arora, R. and Mosch, W. (2011) *High Voltage and Electrical Insulation Engineering*, Wiley-IEEE Press, Chichester.

Cadick, J., Capelli-Schellpfeffer, M. and Neitzel, D. (2005) *Electrical Safety Handbook*, McGraw-Hill, New York.

Degerfalt, M. and Herriman, G. (n.d.) The Homogeneous SA “Substation Automation” Solution, paper by ABB Inc. Canada.

Engmann, G., Fletcher, P., Luigi, G., et al. (2001) The substation design process – An overview, *CIGRE SC23 Colloquium, Puerto Ordaz*.

Matei, G. (2008) New standard solutions for data acquisition in Intelligent Protection Systems, *U.P.B. Sci. Bull., Series C, 70(3)* (available at http://www.scientificbulletin.upb.ro/rev_docs_arhiva/full0377.pdf, accessed May 26, 2015).

McDonald, J.D. (2012) *Electric Power Substations Engineering*. 3rd Edition, CRC Press, Boca Raton.

What about basic Web Functionality?

Some CSS hacks from the Writer's Guide:

The idea of prim Mr. Fleagle plucking his nipple from boneless gums was too much for the class. We burst into gasps of irrepressible snickering. Mr. Fleagle stopped.

“There is nothing funny, boys, about giving suck to a babe. It is the — the very essence of motherhood, don't you see.”

He constantly sprinkled his sentences with “don't you see.” It wasn't a question but an exclamation of mild surprise at our ignorance. “Your pronoun needs an antecedent, don't you see,” he would say, very primly. “The purpose of the Porter's scene, boys, is to provide comic relief from the horror, don't you see.”

Late in the year we tackled the informal essay. “The essay, don't you see, is the...” My mind went numb. Of all forms of writing, none seemed so boring as the essay. Naturally we would have to write informal essays. Mr. Fleagle distributed a homework sheet offering us a choice of topics. None was quite so simpleminded as “What I Did on My Summer Vacation,” but most seemed to be almost as dull. I took the list home and dawdled until the night before the essay was due. Sprawled on the sofa, I finally faced up to the grim task, took the list out of my notebook, and scanned it. The topic on which my eye stopped was “The Art of Eating Spaghetti.”

This title produced an extraordinary sequence of mental images. Surging up out of the depths of memory came a vivid recollection of a night in Belleville when all of us were seated around the supper table — Uncle Allen, my mother, Uncle Charlie,

```
<p id="bed000999"
class="ch2-indent-anofon">&#8222;There is
nothing funny, boys,...<span
class="ch2-line_number">5</span> ..</p>
```

```
<p >He constantly sprinkled ...<span
class="ch2-line_number">6</span>
```

```
.ch2-line_number
```

```
{text-indent: 0em; text-align: right; width: 7%;
margin: 0.4em -9% auto auto; padding: 0em;
font-size: 90%; display: inline-block; float: right;}
```

5
6
7
8

Unicode support

From *Applied English Phonology*

atlas [æʔləs], Atlanta [əʔləntə], he hit me [hihiʔmi], eat well [iʔwɛl],
hot water [hɑʔwɑɾə]. While the glottal stop can replace the /t/ in these

Interactivity can be a challenge

Security concerns and being offline mean that JavaScript is often not supported.

EPUB does not require scripting support, so often publishers do not include it.

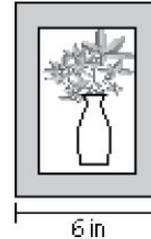
EPUB supports scripting, but is often disabled by Reading Systems before display on platforms that are actually scripting-capable.

From the GMAT Test Prep book:

177. What is the tenths digit of the number d when it is written as a decimal?

1. $d = \frac{54}{25}$

2. $1,000d = 2,160$



178. A framed picture is shown above. The frame, shown shaded, is 6 inches wide and forms a border of uniform width around the picture. What are the dimensions of the viewable portion of the picture?

1. The area of the shaded region is 24 square inches.
2. The frame is 8 inches tall.

5.4 Answer Key

177. D

178. C

179. E

180. D

181. E

182. A

183. D

Implementation issues due to gaps in OWP

EPUB Reading systems need to extend the OWP in order to meet user needs. This can cause implementation problems.

- **Pagination**—reading systems often polyfill pagination using CSS multi-column layout. But this has a side effect of breaking media queries, and interfering with JS execution.
- **Personalization**—reading systems need to implement user font choice, night modes, etc. This can interfere with author CSS, and inappropriately invert image color spaces.
- **Notes.** iBooks nicely pops up footnotes, but without defined behaviors, other reading systems break, and a11y is compromised.

barbed iron from one side of the world to the other.

Good again; but then all confess that somehow whaling

1

See subsequent chapters for something more on this head.

imperial! By

fish."*

has never f

GRAND IM

Roman gen

whale, broug

way from the Syrian coast, were the most conspicuous objects in the cymballed procession.¹

Grant it, since you cite it; but, say what you will, there is no dignity in whaling.

NO DIGNITY IN WHALING? The dignity of our calling the heavens attest. Cetus is a constellation in the South! No more do you bow down your hat in presence of the Czar, and take it off to Queen Victoria. No more! I know a man that, in his lifetime, has taken three hundred and fifty whales. I account that man more honourable than the captain of antiquity who boasted of taking as many walled towns.

Prior Art (aka Opportunities for Reuse)

Components: service worker scope, Web App Manifest scope, EPUB <manifest>

Sequence: link rel=prev/next, EPUB <spine>, HTML <nav> element

Personalization: browser reading modes, ARIA Personalization efforts

Annotations: Web Annotations

Offline: service workers, appcache, local storage

Array of objects: Window.frames

More Prior Art, from XML and HTML

XInclude

```
<book>
  <title>Moby-Dick</title>
  <xi:include href="chapter-01.xml"/>
  <xi:include href="chapter-02.xml"/>
  <xi:include href="chapter-02.xml"/>
</book>
```

HTML Imports

```
<link rel="import" href="c001.html">
<link rel="import" href="c002.html">
<link rel="import" href="c003.html">
```

Recent Efforts

- EPUB BFF (Browser-friendly format)—EPUB, unzipped, with JSON package document describing components, sequence, and metadata
 - IDPF deferred because of pending relationship with W3C
- Radium Web Publication Manifest—similar to BFF
 - Currently a sandbox experiment
- HTML First/EPUB Zero—sequence defined with HTML `<nav>` rather than JSON or XML
 - A thought experiment
- IIIF presentation API—for images, but addresses issues of sequence & metadata.
- Many online journals and magazines present one HTML article at a time from a Table of Contents.

Web Publications

- Address components, sequence, and metadata with JSON manifest.
- Web Publications includes the InfoSet (<https://w3c.github.io/wpub/#infoSet>).
TBD: does this work with Web App Manifest?
- WP must be offline-able: TBD: does this work with Service Workers?
- How does a UA process the publication as a whole?
 - The user should be able to search a whole book, not just the current chapter.
 - Pagination should be done across a whole journal, not just one article (as the publisher sees fit).

Documents are not Applications

- Web of Documents and the Web Operating System.
- Web Apps vs Documents: it's not a problem if an app carries all the code, including a service worker.
- Publishing's central paradigm is the document, and publishers do not want the weight of complex JavaScript in every publication.
- See Lifecycle (<https://w3c.github.io/wpub/#wp-lifecycle>)

Thank you

Tzviya Siegman

Garth Conboy

Dave Cramer

Ivan Herman