The background features a collage of various items: an open book with yellow pages at the top center, a blue book with a globe on its cover to the right, a green book with 'ECONOMICS' written on its spine to the far right, and a hand holding a smartphone in the bottom left corner. The text is overlaid on this collage.

Digital Publishing and the Open Web: The W3C's Digital Publishing Interest Group

CONTEC

Frankfurt a.M., October 8, 2013

Markus Gylling, IDPF, and Ivan Herman,
W3C

These slides will be available on-line:

<http://www.w3.org/2013/Talks/1008-Frankfurt-MGIH/Talk.pdf>



Some words on the International Digital Publishing Forum (IDPF)

<http://www.idpf.org>

International Digital Publishing Forum (IDPF)

- ▶ Not for profit Trade and Standards organization with < 400 members from 40+ countries, working together to advance the Digital Publishing Ecosystem with open, interoperable standards
- ▶ Members include publishers, vendors, libraries, national associations, government and industry organizations
- ▶ Focus is EPUB digital publication format
- ▶ Includes large accessibility contingent, e.g. DAISY Consortium

IDPF activities 2013-14 idpf.org/ongoing

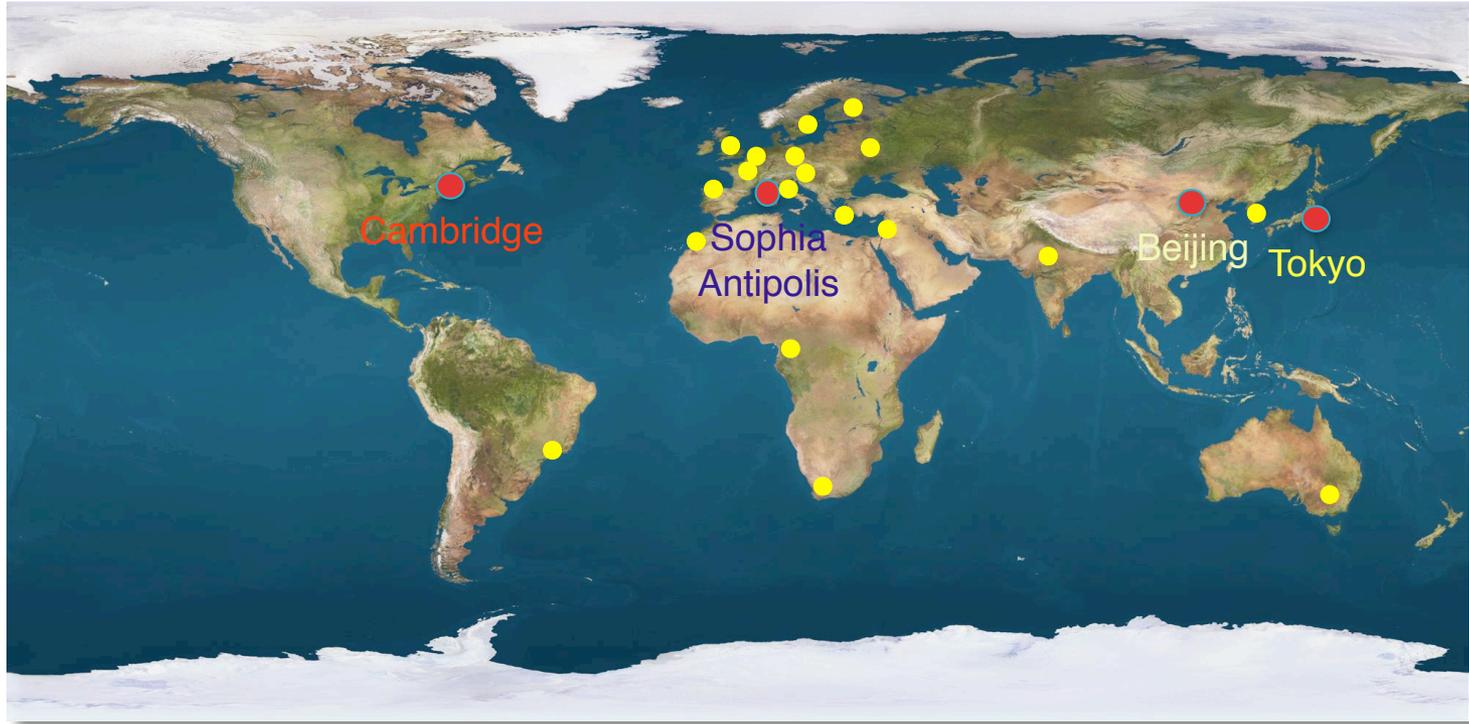
- ▶ EPUB 3.0.1, also standardized as ISO TS
- ▶ Modular Extensions
 - Dictionaries and Glossaries
 - Indexes
 - Interoperable Annotations
 - Multiple Rendition EPUB
 - (dynamic rendition selection & inter-rendition mapping)
 - Profiles for specific domains (e.g. magazines, education)
- ▶ EPUB Reading System test suite
- ▶ *... and W3C collaboration!*

IDPF member open meeting 14:00
Radium Foundation meeting 15:30
Gold 3 room at Marriott

Some words on the World Wide Web Consortium (W3C)

<http://www.w3.org>

W3C is an international community around 4 “hosts”



- ▶ W3C is a non-profit membership organization with more than 370 members from all industries

What W3C does

- ▶ Develops specifications (i.e., Web standards)
 - e.g., HTML, SVG, ...
- ▶ Develops guidelines, best practices
 - e.g., Web Content Accessibility Guidelines, Data Publishing Best Practices, ...
- ▶ Provides diverse forums for discussions
- ▶ Offers a validation suite and service, various courses, documentations

What W3C does not do

- ▶ Major software developments, standard implementation (except for proof of concepts)
- ▶ Conduct independent research (it does what the members want it to do...)
- ▶ Act as a legislative body (e.g., to *require* certain implementations to happen)

What does W3C have to do with publishing?

An ebook is sort-of a website

- ▶ An EPUB file is, technically, a frozen and packaged Web site
 - XHTML 1.1 for EPUB 2
 - XHTML 5 for EPUB 3
- ▶ *An EPUB reader is “just” a very specialized browser!*

Chapter 3. Big Data Tools, Techniques, and Strategies

Designing Great Data Products

By Jeremy Howard, Margit Zwemer, and Mike Loukides

In the past few years, we've seen many data products based on predictive modeling. These products range from weather forecasting to recommendation engines to services that predict airline flight times more accurately than the airlines themselves. But these products are still just making predictions, rather than asking what action they want someone to take as a result of a prediction. Prediction technology can be interesting and mathematically elegant, but we need to take the next step. The technology exists to build data products that can revolutionize entire industries. So, why aren't we building them?

To jump-start this process, we suggest a four-step approach that has already transformed the insurance industry. We call it the *Drivetrain Approach*, inspired by the emerging field of self-driving vehicles. Engineers start by defining a clear *objective*. They want a car to drive safely from point A to point B without human intervention. Great predictive modeling is an important part of the solution, but it no longer stands on its own; as products become more sophisticated, it disappears into the plumbing. Someone using Google's self-driving car is completely unaware of the hundreds (if not thousands) of models and the petabytes of data that make it work. But as data scientists build increasingly sophisticated products, they need a systematic design approach. We don't claim that the Drivetrain Approach is the best or only method; our goal is to start a dialog within the data science and business communities to advance our collective vision.

Objective-based Data Products

We are entering the era of data as drivetrain, where we use data not just to generate more data (in the form of predictions), but use data to produce actionable outcomes. That is the goal of the Drivetrain Approach. The best way to illustrate this process is with a familiar data product: search engines. Back in 1997, AltaVista was king of the algorithmic search world. While their models were good at finding relevant websites, the answer the user was most interested in was often buried on page 100 of the search results. Then, Google came along and transformed online search by beginning with a simple question: What is the user's main objective in typing in a search query?

```
graph LR; A[Defined Objective] --> B[Levers]; B --> C[Data]; C --> D[Models];
```

The four steps in the Drivetrain Approach.

Google realized that the objective was to show the most relevant search result; for other companies, it might be increasing profit, improving the customer experience, finding the best path for a robot, or balancing the load in a data center. Once we have specified the goal, the second step is to specify what inputs of the system we can control, the *levers* we can pull to influence the final outcome. In Google's case, they could control the ranking of the search results. The third step was to

The Web has become an Open Web Platform

- Web pages are more attractive, interactive and intelligent
- HTML5, SVG, or CSS provides cross-browser and cross-device interoperability
- Video, audio, etc., are a first-class citizens
- Data integration has been simplified
- It has tools for social networking (privacy, security, identity)
- Is the most interoperable platform in the industry

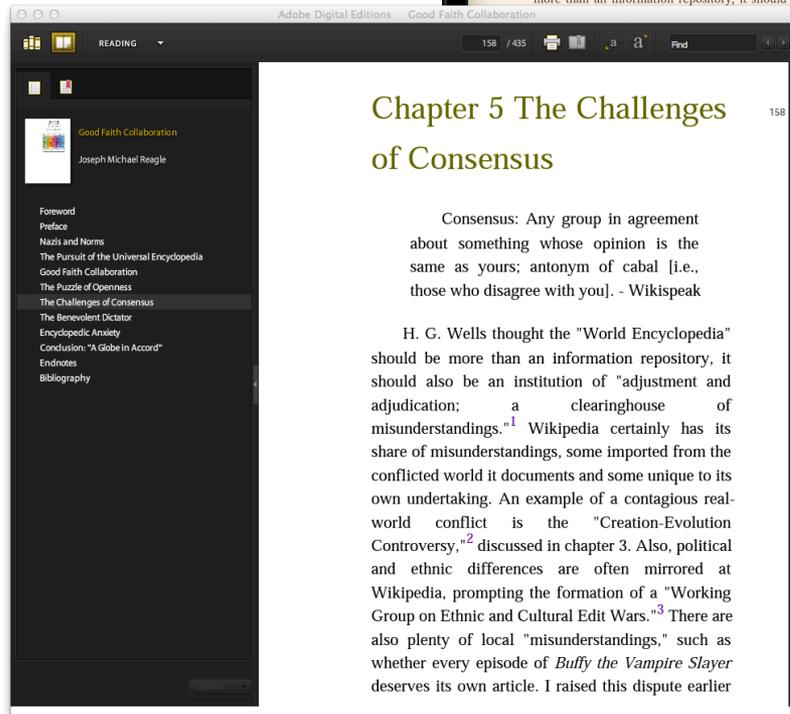
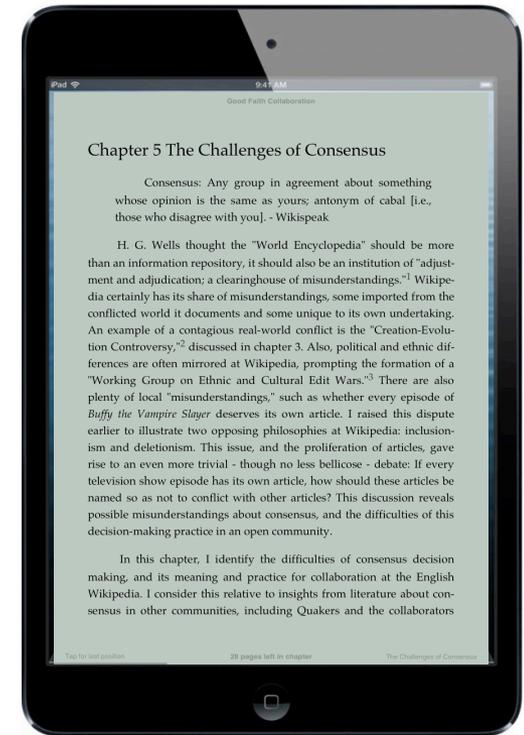
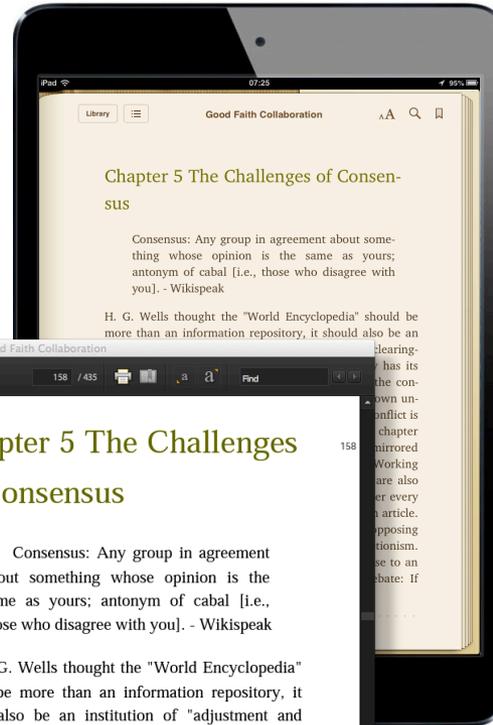
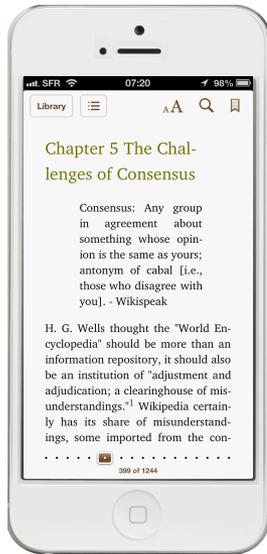


The Open Web Platform has influenced publishing

- ▶ EPUB 3 has picked up on many of the OWP features
 - uses HTML5, i.e., includes audio and video as first class objects
 - SVG comes to the fore
 - CSS 3 instead of older versions of CSS
 - (limited) scripting
 - MathML



And we have some nice results...



And we have some nice results...

And we have some nice results...

GCSE MATHEMATICS

With inbuilt Search Facility

REVISION TEST

Age 15 to 16



GCSE Mathematics Revision Test, is based on our GCSE Mathematics:

- Book 1 of 3 books: Number, Weights and Measures
- Book 2 of 3 books: Algebra
- Book 3 of 3 books: Shape, Space, Statistics and Probability

Included in this book are:

- Polygons, Tessellation, Circle, Locus, Measure, Net, Trapezium Rule, Mid-ordinate Rule, Simpson's Rule, Perimeter, Area, Volume, Bearing, Symmetry, Vector, Mean, Median, Mode, Range, Frequency, Tally Chart, Pie Chart, Cumulative Frequency, Ogive, Probability, Probability Tree and Laws of Probability.

www.mathslearning.com

THE AUTHOR

R.M. O'Toole B.A., M.C., M.S.A., C.I.E.A.

The author is an experienced state examiner, teacher, lecturer, consultant and author in mathematics.

She is an affiliate member of The Chartered Institute of Educational Assessors and a member of The Society of Authors.

First Published in 1996 (Revised in 2012) by:

Mathematics Publishing Company,
'The Cottage',
45, Blackstaff Road,
Clough,
Downpatrick,
Co. Down,
BT30 8SR.

Tel. 028 44 851818 Fax. 028 44 851842

Email address: enquiries@mathslearning.com

COPYRIGHT NOTICE

All rights reserved: no part of this work may be reproduced without either the prior written permission of the Publishers or a licence permitting restricted copying issued by the Copyright Licensing Agency, 90, Tottenham Road, London, W1P 9HE. This book may not be lent, resold, hired out or otherwise disposed of by way of trade in any form of binding or cover other than that in which it is published, without the prior consent of the Mathematics Publishing Company. ©



Is everything perfect and fine?

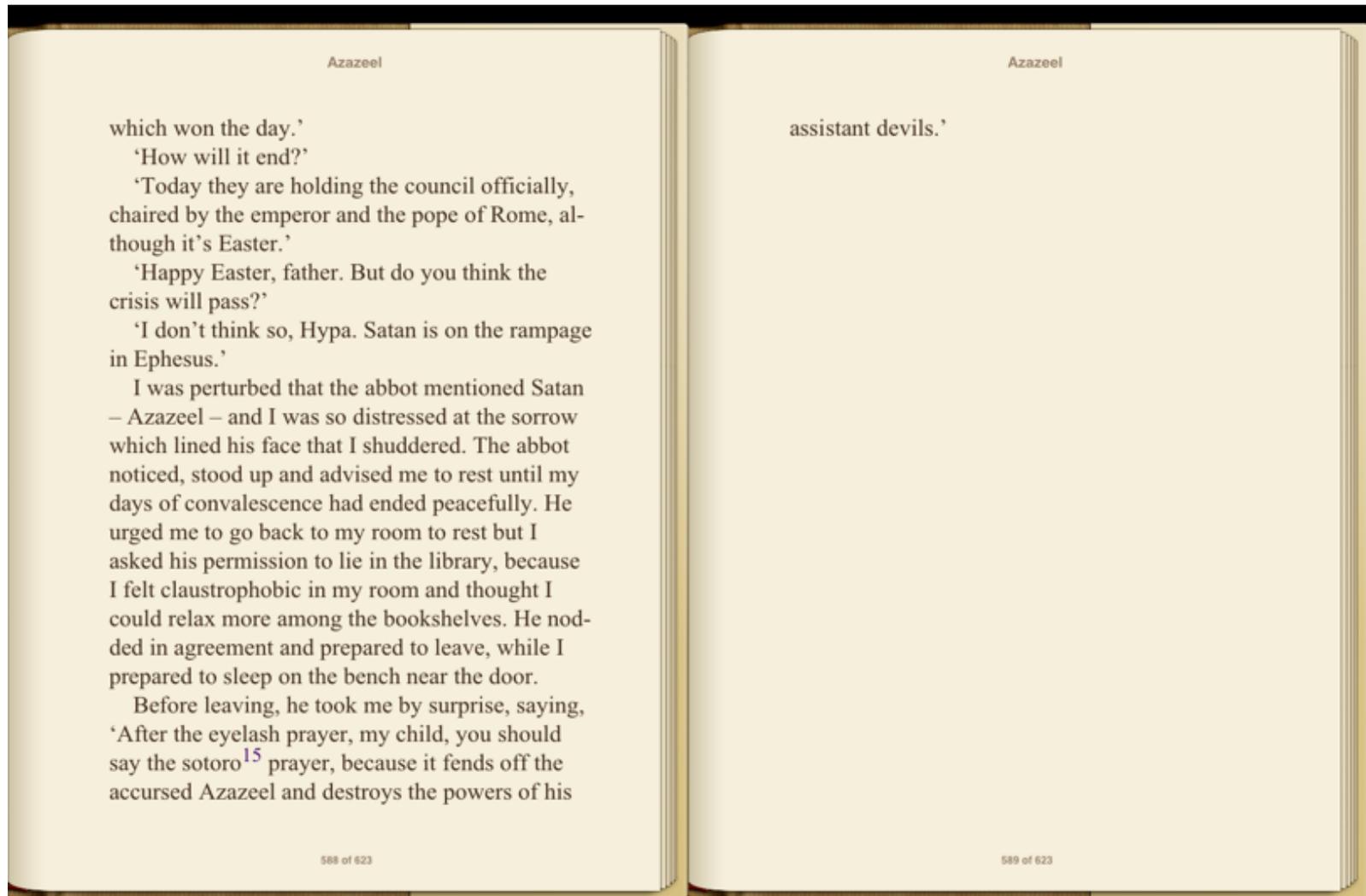
i.e., are publishers and the Web
coexisting perfectly?

No... ☹️

- ▶ Most of the publishing industry players have just been “users” of Web technologies
- ▶ Technology evolution has been driven by “traditional” Web browsing

The specific needs and priorities of publishing may not be reflected in the evolution of the Web!

And this leads to problems...



Missing controls (e.g., in hyphenation)

Chapter 47

WASHINGTON, D.C.

That same afternoon, Senator John Tyler Morgan, Democrat of Alabama, stood in the lobby of the Willard Hotel, yelling at the general manager.

"I have never been refused service in my life! That insufferable man in the elevator had the nerve to tell me he was holding the car for an *important personage*. He told me to get off that car and wait for another car!"

Senator Morgan was so angry that specks of saliva were speckling the lapels of the general manager's morning coat.

"Senator, I am so sorry for the inconvenience—"

"Not an *inconvenience*! It's a goddamned *insult*! Who the hell was he holding the elevator for, the goddamned president of the United States?"

As he roared this question, the great glass doors of the lobby flew open at the hands of two uniformed guards. In walked Theodore Roosevelt.

He took one look at John Tyler Morgan in mid-rampage and the poor little cowering manager. Then Roosevelt thundered, "Unless my eyes de-

bb @

Comp: pe

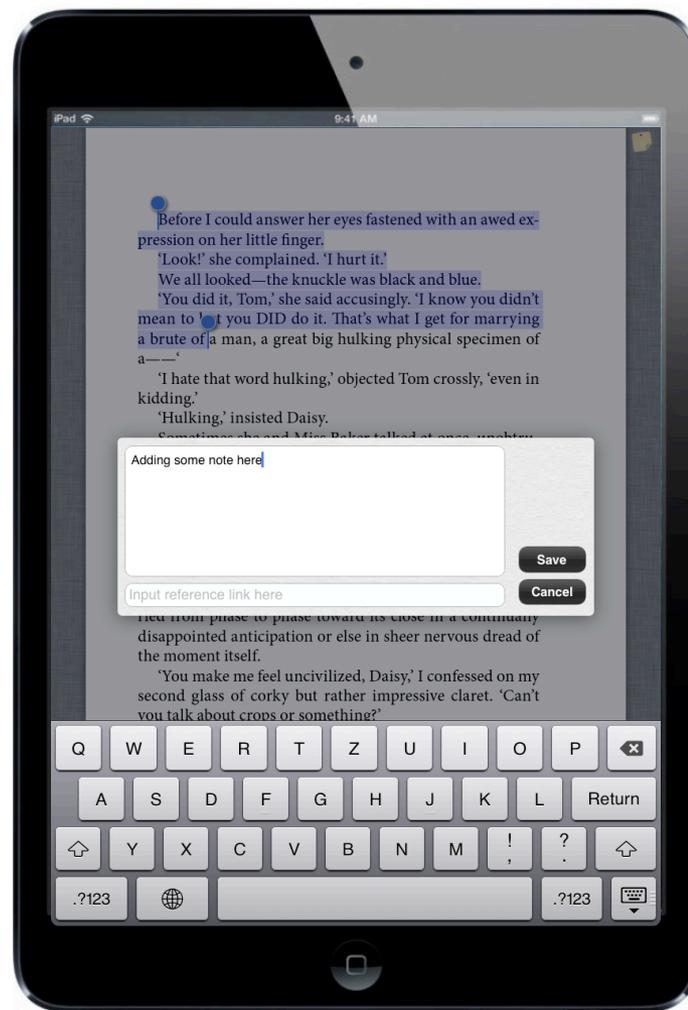
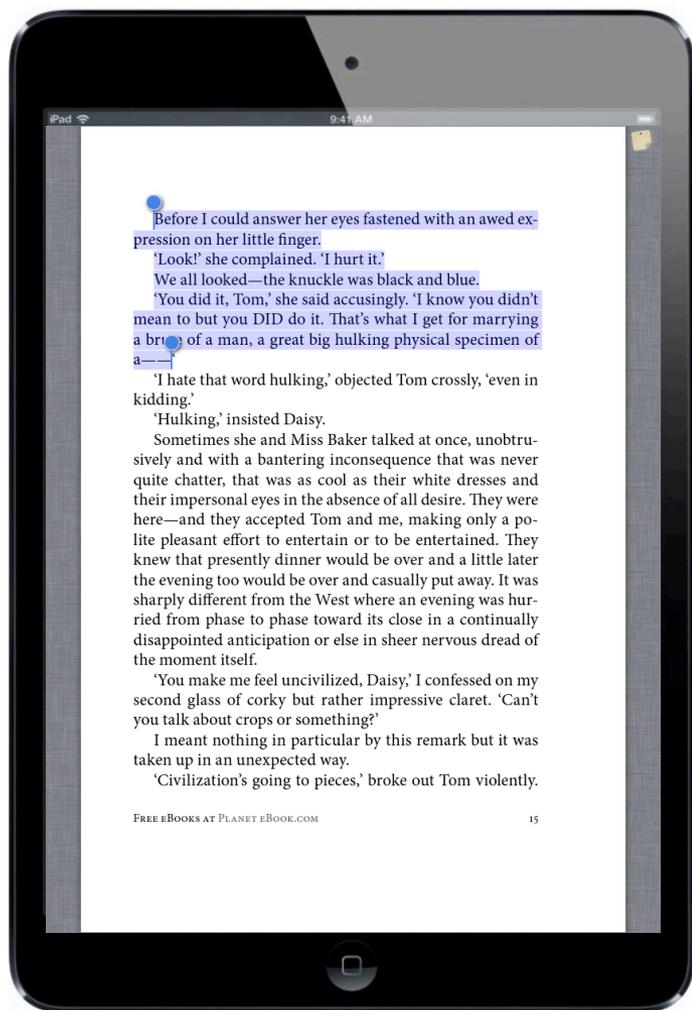
Merriam
Webster,

Cannot
break word

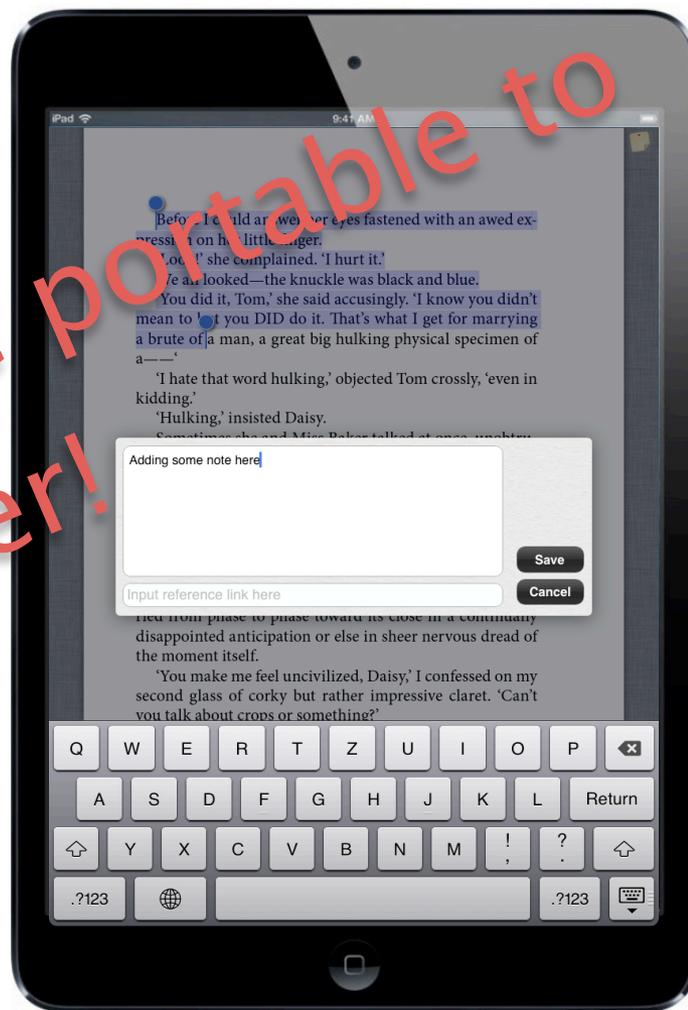
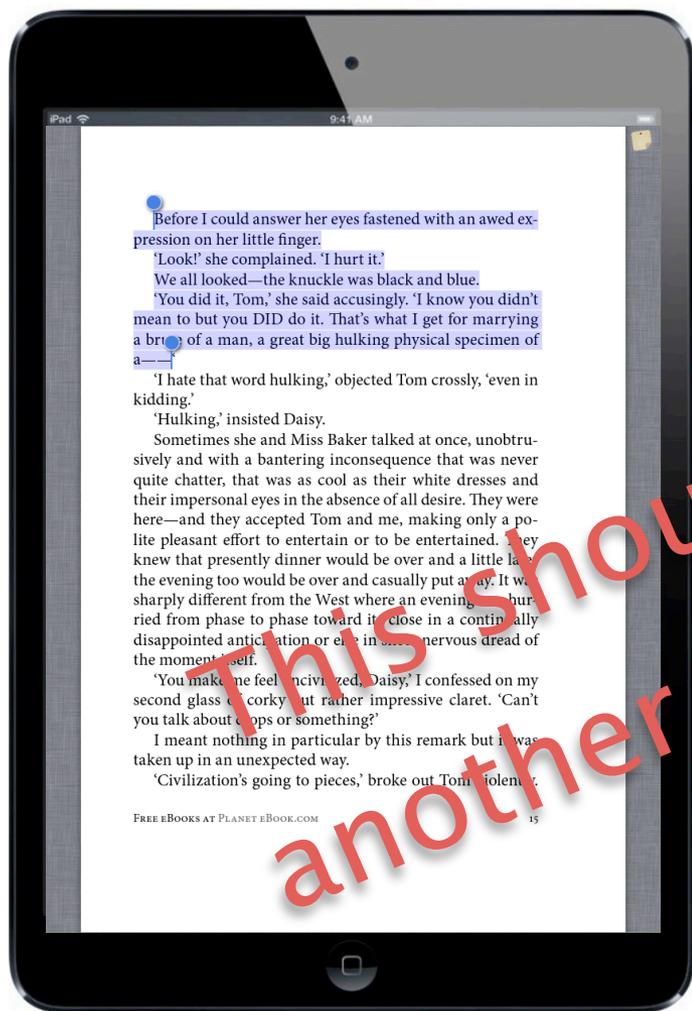
as above.
Must be broken

as:
in·con·ve·nience
can't break
this part.

Non-portable annotations



Non-portable annotations



High quality digital books need

- ▶ Advanced and interactive graphics
- ▶ Internationalization (Ruby, vertical writing)
- ▶ Interoperable annotations and bookmarks
- ▶ MathML
- ▶ Font management
- ▶ Fixed layout
- ▶ Audio, video
- ▶ Local (Web) storage
- ▶ Bridging on-line/off-line
- ▶ Harmonization of metadata
- ▶ Accessibility (e.g., speech/braille publications)
- ▶ ...

High quality digital books need

- ▶ Advanced and interactive graphics
- ▶ Internationalization (Ruby, vertical writing)
- ▶ Interoperable annotations and bookmarks
- ▶ MathML
- ▶ Font management
- ▶ Fixed layout
- ▶ Audio, video
- ▶ Local (Web) storage
- ▶ Bridging on-line/off-line
- ▶ Harmonization of metadata
- ▶ Accessibility (e.g., speech/braille publications)
- ▶ ...

Most of these are being worked on at W3C, but still need extra care and clear requirements!

To repeat...

The technology evolution of the Web has been driven by “traditional” Web browsing

The specific needs and priorities of the publishing industry may not be reflected in the evolution of the Web!

How can we move on?

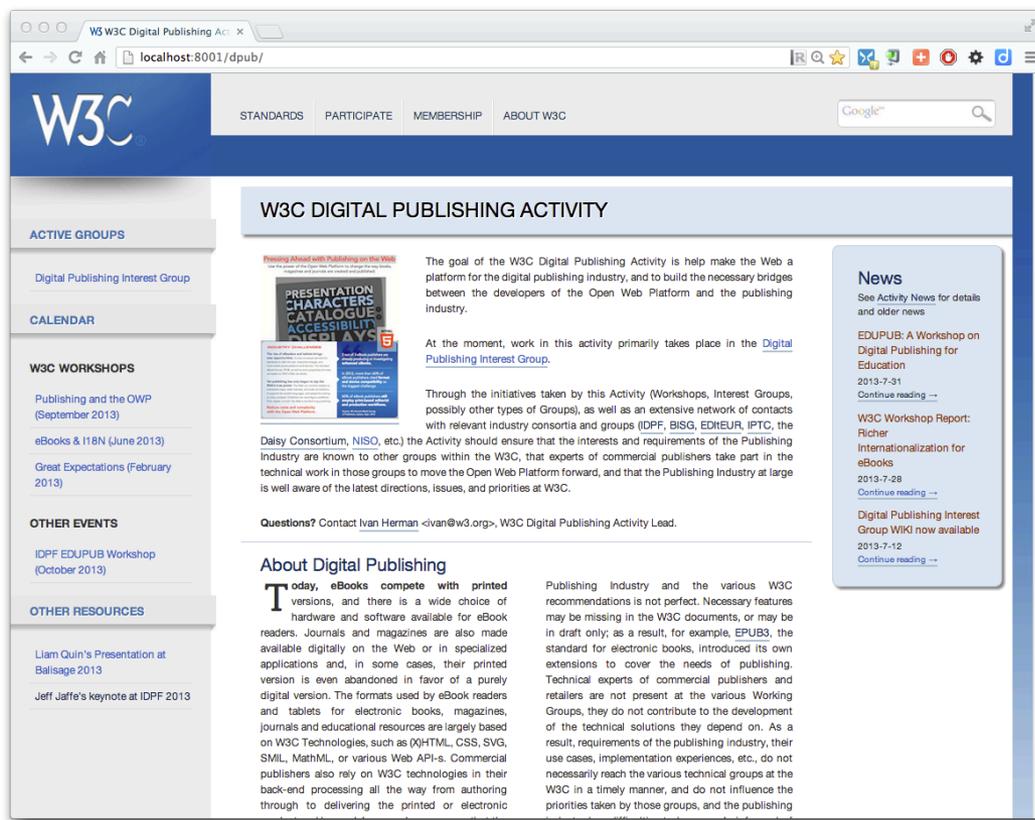
Increasing cooperation between the IDPF and W3C

- ▶ Presence on one another's events
 - Jeff Jaffe (W3C CEO) at IDPF events in NYC and Boston
 - Markus Gylling (IDPF CTO) at W3C's general meeting in Tokyo
- ▶ Workshop organized jointly: “eBooks: Great Expectations”
- ▶ We are also here... :-)

**The objective of the IDPF-W3C relationship is to
*ensure that the EPUB requirements become part
of the central considerations to the future
evolution of the OWP***

Main results so far

- ▶ Creation of a separate Digital Publishing Activity at W3C (June 2013)
 - *Digital Publishing is now a major focus area at W3C*



The screenshot shows a web browser displaying the W3C Digital Publishing Activity website. The browser's address bar shows 'localhost:8001/dpub/'. The website has a blue header with the W3C logo and navigation links for 'STANDARDS', 'PARTICIPATE', 'MEMBERSHIP', and 'ABOUT W3C'. A search bar is located in the top right corner.

The main content area is titled 'W3C DIGITAL PUBLISHING ACTIVITY'. It features a central article with a book cover image titled 'PRESENTATION CHARACTERS CATALOGUE ACCESSIBILITY AND AYS'. The article text discusses the goal of the W3C Digital Publishing Activity to help make the Web a platform for the digital publishing industry and build necessary bridges between developers and the publishing industry. It mentions that work primarily takes place in the Digital Publishing Interest Group and lists various initiatives like Workshops, Interest Groups, and other groups like IDPF, BISG, EDIEUR, IPTC, the Daisy Consortium, and NISO.

On the left side, there are several sidebar sections: 'ACTIVE GROUPS' (Digital Publishing Interest Group), 'CALENDAR', 'W3C WORKSHOPS' (Publishing and the OWP, eBooks & I18N, Great Expectations), 'OTHER EVENTS' (IDPF EDUPUB Workshop), and 'OTHER RESOURCES' (Liam Quin's Presentation, Jeff Jaffe's keynote).

On the right side, there is a 'News' section with two entries: 'EDUPUB: A Workshop on Digital Publishing for Education' and 'W3C Workshop Report: Richer Internationalization for eBooks'. Both entries include dates and 'Continue reading' links.

At the bottom right, there is a logo for '<idpf> W3C'.

New Digital Publishing Activity and Interest Group at W3C

- ▶ Co-chaired by Markus Gylling (IDPF) and Madi Salomon (Pearson)

“The **mission** of the Digital Publishing Interest Group, [...] **a forum for experts in the digital publishing ecosystem** [...] for technical discussions, gathering use cases and to **better align existing formats and technologies (e.g., EPUB) with the broader Open Web Platform.**”

The W3C Digital Publishing Interest Group current & projected focus areas

http://www.w3.org/dpub/IG/wiki/Main_Page

1. Styling and Layout

- ▶ i18n (vertical writing, ruby et. al.)
- ▶ advanced font handling
- ▶ personalization *with precision*
- ▶ browser-native pagination
- ▶ high-design adaptive layout (e.g. regions, exclusions, page templates)

of our papers had been picked up by the AP and had spread worldwide. We did interviews for newspapers and radio stations all across the country and at least one radio station in Canada.

With all the exposure it wasn't long before we were made aware that the record of 120 hours had been broken already. A few emails and phone calls later and we found out that two State Farm agents from Texas held the last record certified by Guinness of 118 hours. That was quite a downer. We had followers counting on us to break the record and we had figured everything on a faulty premise.

We resolved to just keep at it and hope for the best. We received lots of encouragement from friends, family and fans and we couldn't let them down. After

“To anyone we inspired, thanks.”

many miles and a few adventures (which can be found here) we finished our crazy trip at four corners. We beat our expectations by a long shot, hitting all 48 contiguous states in 106 hours and 43 minutes. Even though we had five more hours of driving to get home, we were ecstatic.

After we wrapped up we found out how many people had heard about our Great American

Road Trip. We had as many as 25,000 hits on this site in a day from dozens of countries from every continent except Antarctica. We were mentioned in magazines ranging from a Christian Magazine to Car and Driver. We continued to do interviews and we found news articles in Chinese, German, and many other languages. We were astounded.

So many people connected with our trip in so many ways and we have been an inspiration to many to hit the road or follow their dreams. Some of those we inspired were inspired to beat our record and accomplished just that. The trip was great, the sense of accomplishment was wonderful but knowing we might have made a difference in someone's day, let alone life, is better than any record.

To anyone we inspired, to all those that enjoyed our journey, to our families and to our friends, thank you.



of our
local papers

had been picked up by the AP and had spread worldwide. We did interviews for newspapers and radio stations all across the country and at least one radio station in Canada.

With all the exposure it wasn't long before we were made aware that the record of 120 hours had been broken already. A few emails and phone calls later and we found out that two State Farm agents from Texas held the last record certified by Guinness of 118 hours. That was quite a downer. We had followers counting on us to break the record and we had figured everything on a faulty premise.

We resolved to just keep at it and hope for the best. We received lots of encouragement from friends, family and fans and we couldn't let them down. After many miles and a few adventures (which can be found here) we

finished our crazy trip at four corners. We beat our expectations by a long shot, hitting all 48 contiguous states in 106 hours and 43 minutes. Even though we had five more hours of driving to get home, we were ecstatic.

After we wrapped up we found out how many people had heard about our Great American Road Trip. We had as many as 25,000 hits on this site in a day from dozens of countries from every continent except Antarctica. We were mentioned in magazines ranging from a Christian Magazine to Car and Driver. We continued to do interviews and we found news articles in Chinese, German, and many other languages. We were astounded.

So many people connected with our trip in so many ways and we have been an inspiration to many to hit the road or follow their dreams. Some of those we inspired were inspired to beat our

record and accomplished just that. The trip was great, the sense of accomplishment was wonderful but knowing we might have made a difference in someone's day, let alone

“To anyone we inspired, thanks.”

life, is better than any record.

To anyone we inspired, to all those that enjoyed our journey, to our families and to our friends, thank you.



2. Domain-specific content & behaviors

▶ Education

- scientific notation
- infographics
- embedded assessments (protocols, APIs)
- learning objects: metadata and remixing
 - <http://www.imslobal.org>
 - <http://www.lrmi.net>

▶ Comics/Manga

▶ Magazines

3. Accessibility and Personalization

Accessibility is just an extreme case of personalization

- ▶ Integration of Web accessibility principles in ebooks
 - making sense of dynamic content
 - structural semantic vocabularies (for education and more)
- ▶ Speech, braille and longform publications

4. Internationalization

Japanese writing mode
(written using a top-to-bottom inline direction with a leftward (right-to-left) block flow direction)

Ruby
(typically used in East Asian documents to indicate pronunciation or to provide a short annotation)

portrait of the author

ebook title

草枕



tate-cyu-yoko
(western languages mixed in Japanese)

夏目漱石

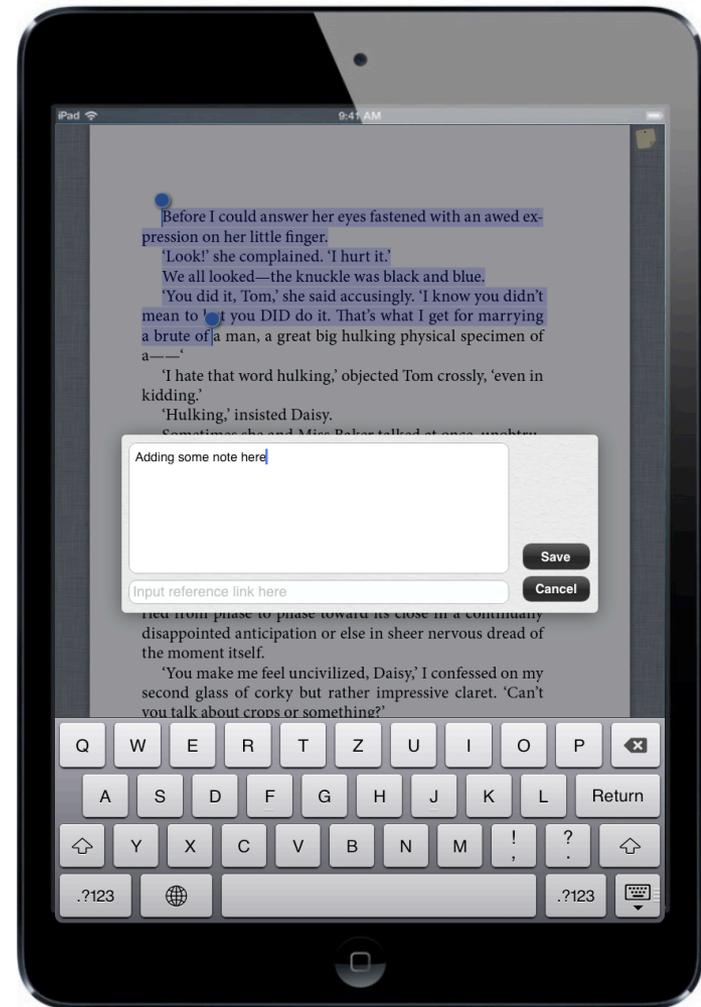
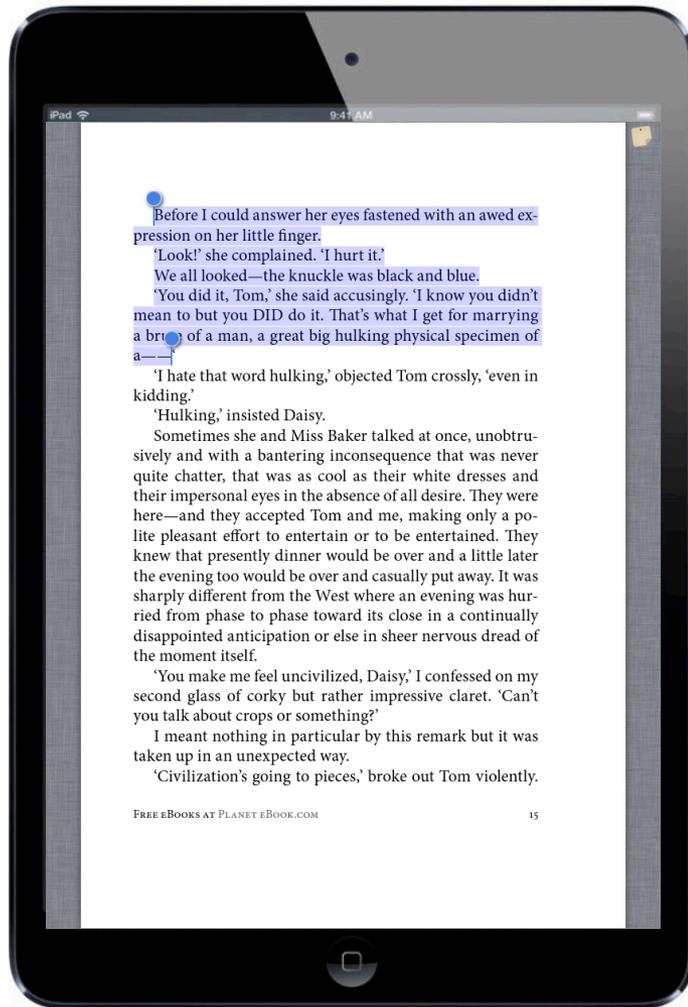
Author

line-break
(line break never occurs just before the Japanese period - ".")

山路を登りながら、こう考えた。
 智に働けば角が立つ。情に惚れば
 越す事のならぬ世に住みにくければ、
 住みにくさが高ると、安い所へ引き越したくなる。ど
 へ越して、住みにくいと悟った時、詩が生れて、画が出来
 人の世を作ったものは神でもなければ鬼でもない。やは
 向う三軒両隣りにちらちらするただの人である。ただの人
 が作った人の世に住みにくいからとて、越す国はあるまい。
 あれば人でなしの国へ行くばかりだ。人でなしの国は人の
 世よりもなお住みにくがる。
 越す事のならぬ世に住みにくければ、住みにくい所をどれ
 ほどか、寛容で、東の間の命を、東の間でも住みよくせねば
 ならぬ。ここに詩人という天職が出来て、ここに画家という
 使命が降る。あらゆる芸術の士は人の世を長閑にし、人の心
 を豊かにするが故に尊とい。
 住みにくき世から、住みにくき煩いを引き抜いて、あり
 がたい世界をまのあたりに写すのが詩である、画である。あ
 るは音楽と彫刻である。こまかに云えば写さないでもよい。
 ただまのあたりに見れば、そこに詩も生き、歌も湧く。着想
 を紙に落さぬとも、瑣屑の音は胸裏に起る。丹青は画架に
 向って塗抹せんでも五彩の絢爛は自から心眼に映る。ただ
 おりがまじ世を、かく観じ得て、筆を動かすの力メラニ

夏目漱石 (Natsume Soseki) 慶
 応三年一月廿五日—大正九年十二月廿日
 は、日本の小説家、評論家、翻訳家、英文学
 者。本名、夏目 金之助。『吾輩は
 猫である』『こゝろ』などの作品
 で広く知られる。江戸の牛込馬場下
 横町出身。俳号は愚陀仏。

5. Social Reading and Annotations



6. Packaging and Distribution

Bridging online and offline/portable

- ▶ online and contained/portable as states, not manifestations
- ▶ discoverability
- ▶ direct vs abstract addressing
 - inter-publication linking (DOI, XRI et. al.)



Thank you for your attention!

EASY NOT IDLE.