

# (Packaged) Web Publications

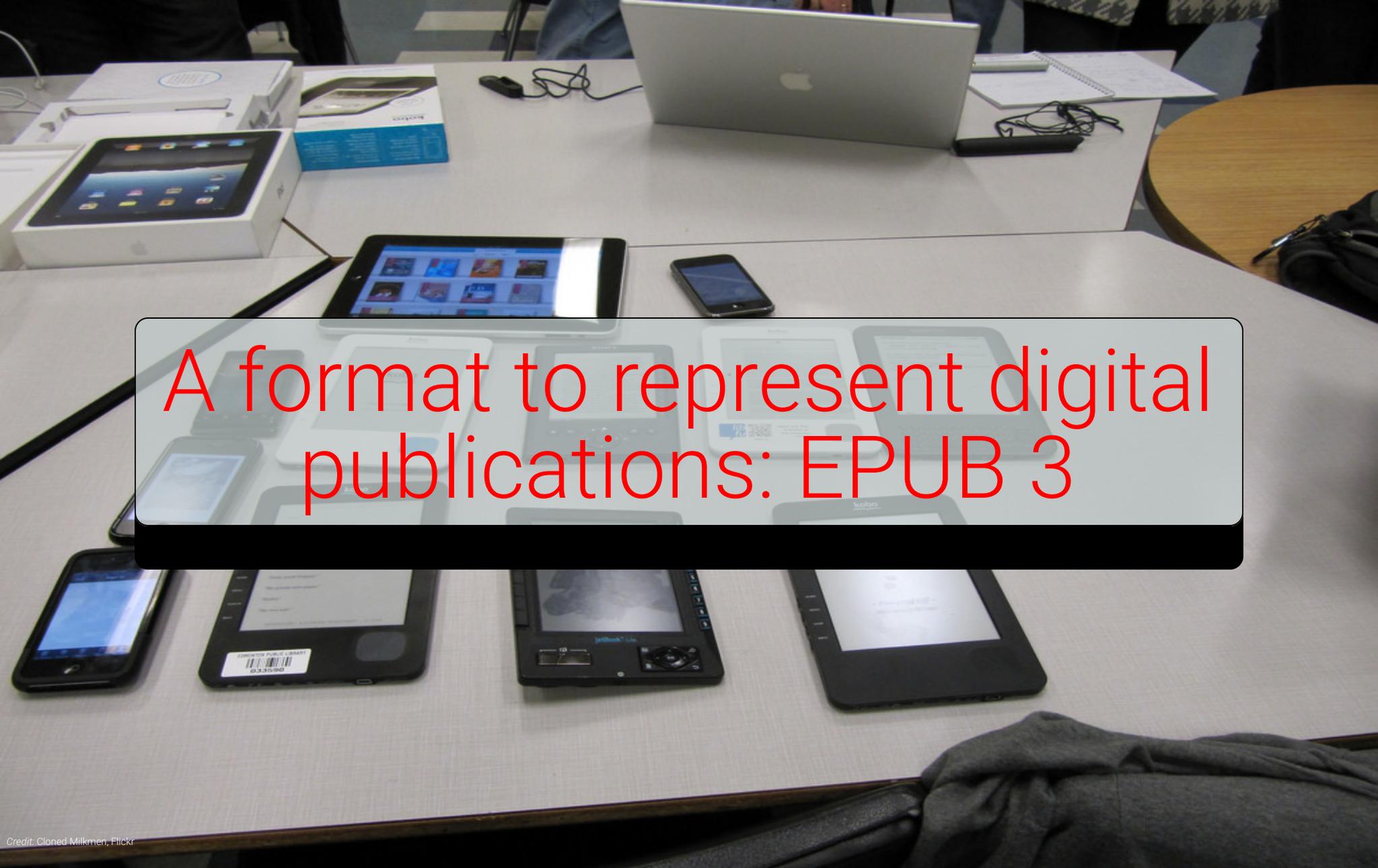
Ivan Herman, W3C  
Semantic Web in Libraries (SWIB16), Bonn, Germany  
2016-11-29

# These Slides are Available on the Web

---

See: <https://w3c.github.io/dpub/2016/SWIB/>

(Slides are in HTML)

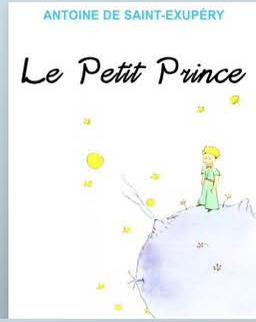
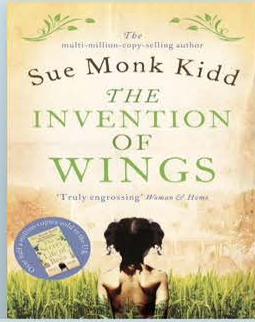
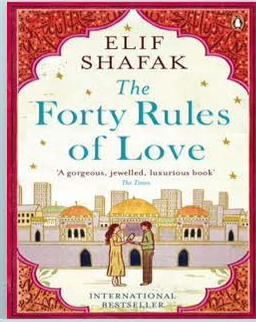
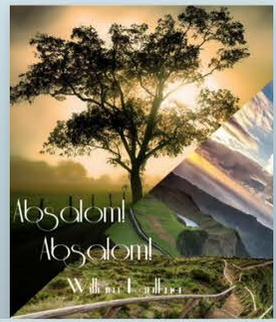
A photograph of a table with various electronic devices including iPads, iPhones, and e-readers. In the background, there are boxes for iPads and iPhones, a laptop, and some papers. The foreground shows several devices: an iPad displaying a grid of app icons, an iPhone, and several e-readers. One e-reader has a library barcode sticker. A semi-transparent white box with a black border and a black drop shadow is centered over the image, containing red text.

# A format to represent digital publications: EPUB 3

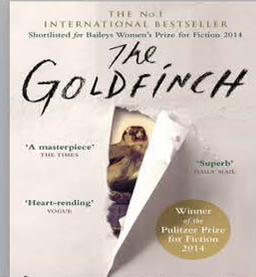
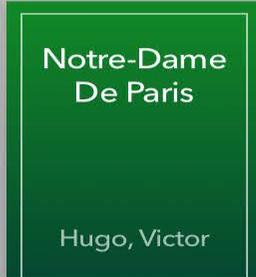
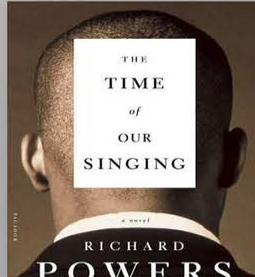
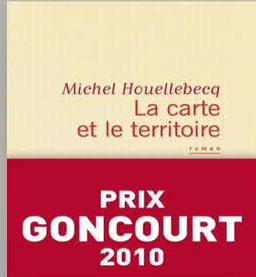
# EPUB 3 is a very mature specification

---

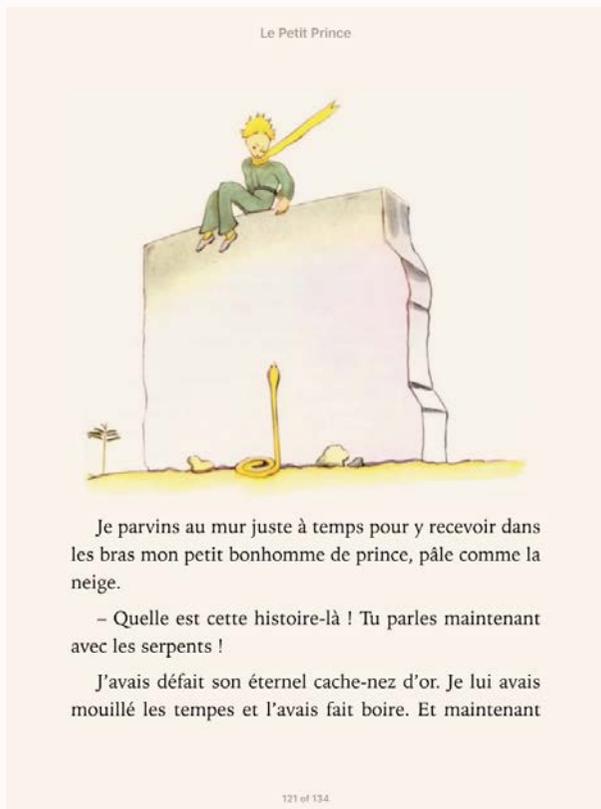
- Widely used format for digital publication
- A truly portable and non-proprietary format
- The standard format for digital books...
- ...but can also be used for magazines, journal articles, documentation, etc.



A wide variety of books have been created



# Books with *lots* of illustrations...



Credit: Antoine de Sain-Exupéry: "Le Petit Prince", Ebooks libres et gratuits

Winnie-the-Pooh

I'll go and tell them.' And he stumped off. The first person he met was Rabbit. 'Hallo, Rabbit,' he said, 'is that you?' 'Let's pretend it isn't,' said Rabbit, 'and see what happens.'



'I've got a message for you.' 'I'll give it to him.' 'We're all going on an Expedition with Christopher Robin!' 'What is it when we're on it?' 'A sort of boat, I think,' said Pooh. 'Oh! that sort.' 'Yes. And we're going to discover a Pole or something. Or was it a Mole? Anyhow we're going to discover it.' 'We are, are we?' said Rabbit. 'Yes. And we've got to bring Pro-things to eat with us. In case we want to eat them. Now I'm going down to Piglet's. Tell Kanga, will you?' He left Rabbit and hurried down to Piglet's house. The

111 of 156

Credit: A.A. Milne: "Winnie-the-P ooh", Egmont UK Ltd.

# Scientific presentations

iPad 15:35 95%

**Content**

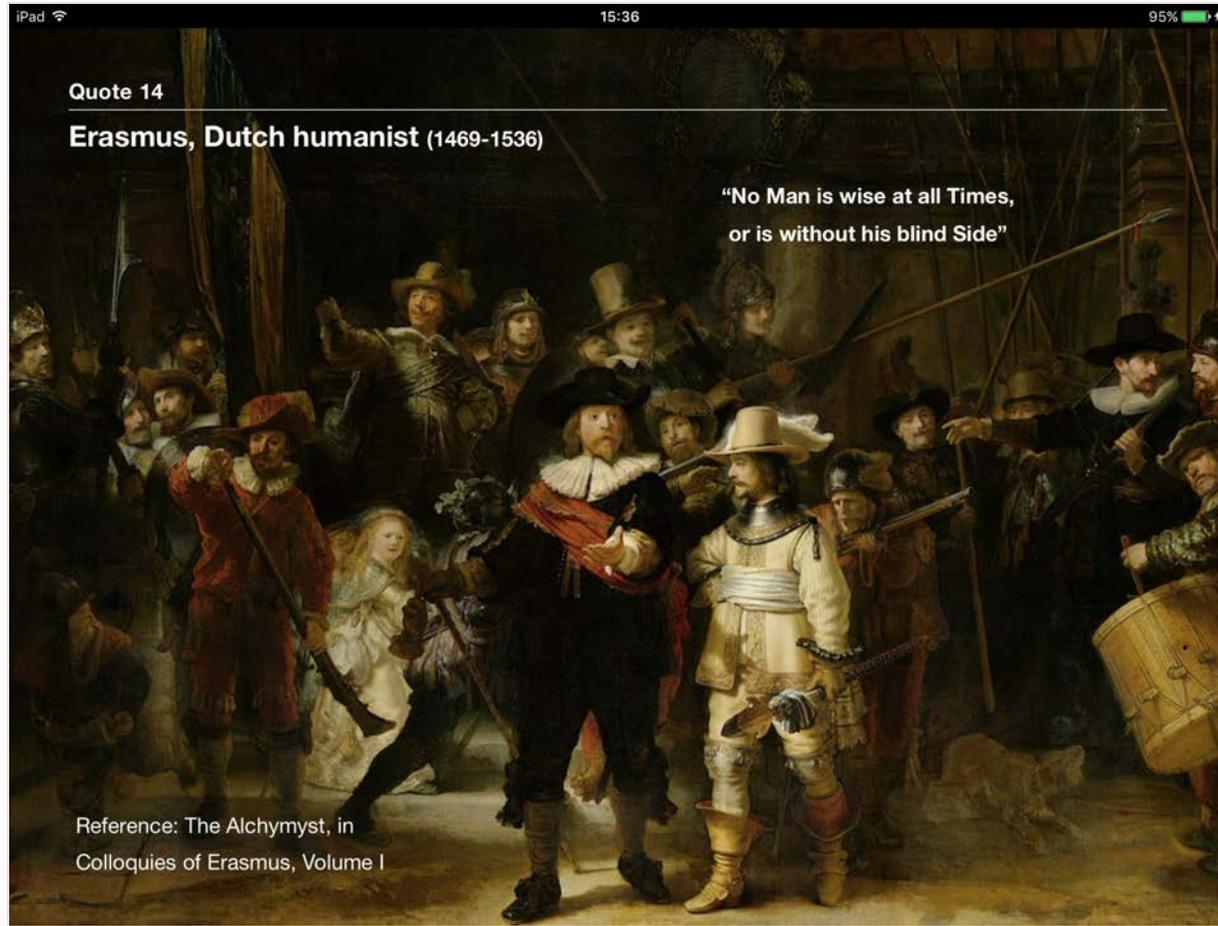
1. Integration by Parts
2. Trigonometric Integrals
3. Trigonometric Substitution
4. Integration of Rational Functions
5. Strategy for Integration
6. Approximate Integration
7. Improper Integrals

**Strategy for Integration**

- Integration need to be memorized:  
 $\int x^n dx$   $\int \frac{1}{x} dx$   $\int a^x dx$   $\int \sin x dx$   $\int \cos x dx$   
 $\int \sec^2 x dx$   $\int \csc^2 x dx$   $\int \tan x \sec x dx$   $\int \cot x \csc x dx$   
 $\int \frac{1}{x^2 - a^2} dx$   $\int \frac{1}{x^2 + a^2} dx$   $\int \frac{1}{\sqrt{x^2 + a^2}} dx$   $\int \frac{1}{\sqrt{a^2 - x^2}} dx$
- Integration need to know how to derive:  
 $\int \tan x dx$   $\int \cot x dx$   $\int \sec x dx$   $\int \csc x dx$

10

# Art books





# Technical books

## CSS Secrets

```
font-size: 125%; /* Assuming a 16px parent font size */
line-height: 1.5;
```

Now if I change the parent font size, the button will instantly become bigger. However, it will look quite different (Figure 1.5), because all other effects were designed for a smaller button and did not scale. We can make all the other effects scalable as well, by specifying any lengths in `ems`, so that they all depend on the font size. This way, we can control the size of the button in one place:

Here we wanted our font size and measurements to be relative to the parent font size, so we used `ems`. In some cases, you want them to be relative to the **root font size** (i.e., the font size of `<html>`), and `ems` result in complex calculations. In that case, you can use the `rem` unit. Relativity is an important feature in CSS, but you do have to think about what things should be relative to.

```
padding: .3em .8em;
border: 1px solid #446d88;
background: #58a linear-gradient(#77a0bb, #58a);
border-radius: .2em;
box-shadow: 0 .05em .25em gray;
color: white;
text-shadow: 0 -.05em .05em #335166;
font-size: 125%;
line-height: 1.5;
```

Yes!

FIGURE 1.5 Enlarging the font size breaks other effects in our button (corner rounding being the most noticeable), as they are specified using absolute lengths

Now our larger button looks much more like a scaled version of the original (Figure 1.6). Notice that we still left some lengths as absolute values. It's a judgment call which effects should scale with the button and which ones should stay the same. In this case, we wanted our border thickness to stay 1px

Credit: Lea Verou: "CSS Secrets", O'Reilly

## Functional JavaScript

information, and warnings can thus be modified entirely:

```
function note() {}
function warn(str) {
  alert("That doesn't look like a valid age");
}

parseAge("frob");
// (alert box) That doesn't look like a valid age
//=> 0
```

Therefore, because the behavior is contained within a single function, the function can be replaced by new functions providing similar behavior or outright different behaviors altogether (Abelson and Sussman 1996).

## Encapsulation and Hiding

Over the years, we've been taught that a cornerstone of object-oriented programming is *encapsulation*. The term encapsulation in reference to object-oriented programming refers to a way of packaging certain pieces of data with the very operations that manipulate them, as seen in Figure 1-8.

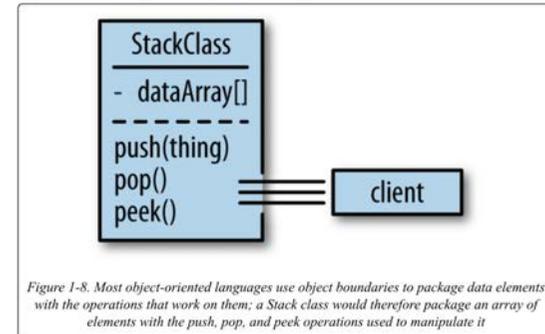


Figure 1-8. Most object-oriented languages use object boundaries to package data elements with the operations that work on them; a Stack class would therefore package an array of elements with the push, pop, and peek operations used to manipulate it

Credit: Michael Fogus: "Functional Programming", O'Reilly

# Books with different character sets and writing directions

- महाभारत
- 42 विविधं संहिता जजानं दीपयन्ति मनीषिणः  
वयाख्यातुं कुशलाः के चिद गन्धं धारयितुं परे
- 43 तपसा ब्रह्मचर्येण वयस्य वेदं सनातनम  
इतिहासम इमं चक्रे पुण्यं सत्यवती सुतः
- 44 पराशरामजो विद्वान् ब्रह्मर्षिः संशितव्रतः  
मातुर नियोगाद् धर्मिणा गाङ्गेयस्य च धीमतः
- 45 कषेत्रे विचित्रवीर्यस्य कृष्णद्वैपायनः पुरा  
तरीन अग्नीन् इव कोरव्याञ्जनयाम आस वीर्यवान्
- 46 उन्वाद्य धृतराष्ट्रं च पाण्डुं विदुरम एव च  
जगाम तपसे धीमान् पुनर एवाश्रमं परति
- 47 तेषु जातेषु वृद्धेषु गतेषु परमां गतिम्  
अब्रवीद् भारतं लोके मानुषे ऽसमिन् महान् ऋषिः
- 48 जनमेजयेन पृष्टः सन् बराह्मणैश्च च सहस्वशः  
शशास शिष्यम् आसीनं वैशम्पायनम् अन्तिके
- 49 स सदस्यैः सहासीनः शरावयाम आस भारतम्  
कर्मान्तरेषु यज्ञस्य चोद्यमानः पुनः पुनः
- 50 विस्तरं कुरुवंशस्य गान्धार्या धर्मशीलताम्  
कषतुः परसां भृतिं कुन्त्याः सम्यग् दवैपायनो ऽब्रवीत्
- 51 वासुदेवस्य माहात्म्यं पाण्डवानां च सत्यताम्  
दुर्वृत्तं धार्तराष्ट्राणाम् उक्तवान् भगवान् ऋषिः
- 52 चतुर्विंशतिसाहस्रीं चक्रे भारतं संहिताम्  
उपाख्यामेर विना तावद् भारतं परोच्यते बुधैः
- 53 ततो ऽधर्धशतं भूयः संक्षेपं कृतवान् ऋषिः  
अनुक्रमणिम् अध्यायं वृत्तान्तानां सर्पवर्णाम्
- 54 इदं दवैपायनः पूर्व पुत्रम् अध्यापयच छुक्रम  
ततो ऽनयेभ्यो ऽनुरूभ्यः शिष्येभ्यः परददौ परमुः
- 55 नारदो ऽशरावयद् देवान् असितो देवलः पितृन्  
गन्धर्वयक्षरक्षांसि शरावयाम् आस ये शुकः
- 56 दुर्योधनो मन्सुमयो महाद्रुमः सकन्धः कर्णः शकुनिस तस्य शाखाः  
दुःशासनः पुष्यफले समृद्धेः मूलं राजा धृतराष्ट्रो ऽमनीषी
- 57 युधिष्ठितो धर्ममयो महाद्रुमः सकन्धो ऽरजुनो भीमसेनो ऽसय शाखाः  
माद्री सुतो पुष्यफले समृद्धेः मूलं कृष्णो बरह्म च बराह्मणाश्च
- 58 पाण्डुर जित्वा बहून् देशान् युधा विक्रमणेन च  
अरण्ये मृगया शीलौ नयवसत् सजनस तदा
- 59 मृगव्यवायं निधने कृच्छ्रां पराप स आपदम्  
जन्मप्रभृति पाथानां तत्राचार विधिक्रमः

מפליגים בישראל



לפני מספר חודשים הרס וירוס אימתני את מרבית מחשבי העולם האירוע הזה זכה לשם "קריסת המחשבים הגדולה" נחשב לאירוע הטרוור הגדול ביותר מאז 11 בספטמבר 2001 ולמרבית הפלא קרה בדיוק באותו תאריך בשנת 2014.

תוך כדי שהוא עסוק בהרהוריו, צילצל הטלפון. 'אולי היא כבר מתקשרת?' חשב אבישי תוך שהוא מוציא את הטלפון מכיסו.

"שלום, אבישי! מדברת אלכס, זוכר אותי?"

לקח לו שנייה אחת להוכיח. אלכס הייתה אשת המכירות בחברה ממנה רכש את היאכטה שלו לפני כשנתיים. עד לפני שנה היו מתכתבים ביניהם. אבל שניהם לא היו ממש טובים במשימת קשר מרוחק וכך נתמסס לו הקשר.

"כן, בוודאי שאני זוכר, מה את עושה בארץ?" שאל אבישי באנגלית.

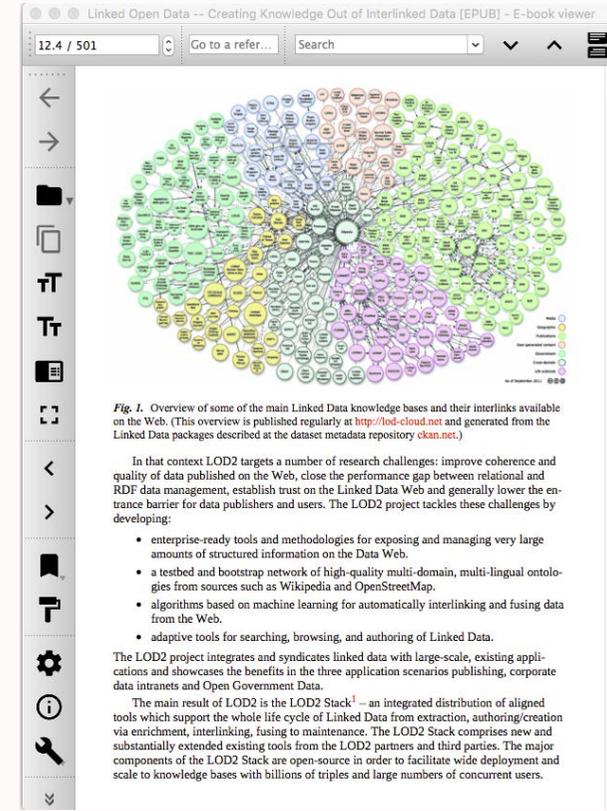
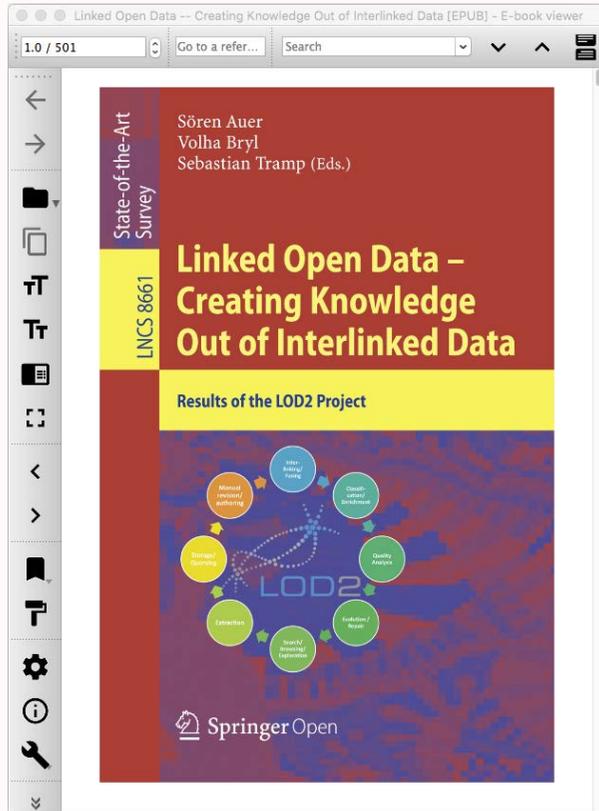
"אני בישראל לכמה הרצאות ופגישות עם לקוחות, נוכרתי שכתבת לי על היאכטה במרינה הרצליה אז הלכתי היום לננהלת המרינה לחפש אותך." ענתה אלכס באנגלית המדויקת שלה כמעט ללא מבטא צרפתי.

אבישי שמח שהיא זכרה אותו וזכרה אותה לחפש אותו.

"טעולה, את מוזמנת לבוא לאכטה, איפה את מאחסנת?" שאל אבישי.

It is not only for books! It can  
be...

# ...conference proceedings



# ...journals or magazines (articles)

The Journal of Electronic Publishing: 20 Years of Electronic Publishing



20 Years of Electronic Publishing

Edited By Maria Bonn and Jonathan McGlone

The Journal of Electronic Publishing: 20 Years of Electronic Publishing

## Contents

[Economics of Electronic Publishing](#)

Colin Day

Volume 1, Issue 1&2, January 1995

[The Role of Responsive Pricing in the Internet](#)

Jeffrey K. Mackie-Mason, Liam Murphy, and John Murphy

Volume 1, Issue 1&2, January 1995

[Waiting for Thomas Kuhn: First Monday and the Evolution of Electronic Journals](#)

Edward J. Valauskas

Volume 3, Issue 1, September 1997

[The Online Newspaper: A Postmodern Medium](#)

Shayla Thiel

Volume 4, Issue 1, September 1998

[Copyright in a Time of Change](#)

William S. Strong

Volume 4, Issue 3, March 1999

[Reprint: Five Years and Counting](#)

John Labovitz

Volume 5, Issue 2: What is Past is Prologue, Dec., 1999

[The Impact of the Internet on Teaching and Practicing Journalism](#)

Joanne Teoh Kheng Yau and Suliman Hawamdeh

Volume 7, Issue 1: Taking License, August, 2001

[White Paper: The Deep Web: Surfacing Hidden Value](#)

Michael K. Bergman

Volume 7, Issue 1: Taking License, August, 2001

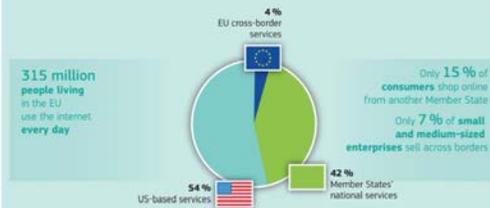
[What Are The Alternatives To Peer Review? Quality Control in Scholarly](#)

# ...official reports of all kinds

European Commission

that sell their products digitally also complain that delivery costs are a problem. The Commission is promoting *affordable, high-quality, cross-border delivery services* through the digital single market strategy, so that businesses can bring their products to consumers in a way that benefits all those involved in the transaction. To find out exactly what citizens and businesses need and want the Commission launched a *consultation* on this issue in May.

## E-COMMERCE: HUGE POTENTIAL BUT LARGELY UNUSED AT EU LEVEL



Source: Joint Research Centre/Institute for Prospective Technological Studies, digital economy working paper, 2015.

### Tackling geo-blocking

Many citizens who shop online have fallen victim to geo-blocking, the commercial practice whereby online sellers either deny consumers access to a website based on their location or reroute them to a local store with different prices. For example, a car rental customer in one Member State may end up paying more for an identical car rental than customers from another Member State. In September the Commission launched a *public consultation* to understand the needs of consumers, in order to help it draft legislative proposals to end this practice, if unjustified.

### Modernising copyright law

Today, EU residents travelling within the European Union may be cut off from online services providing films, sports broadcasts, music, e-books and games that they have

7 pages left in chapter

Maximising growth potential

Credit: European Commission – General Report 2015

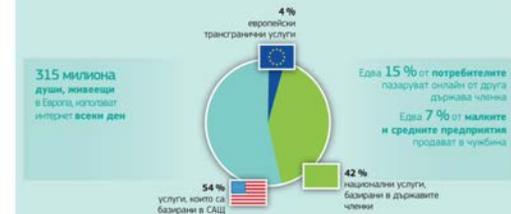
Европейска комисия

дружествата в ЕС и да гарантира, че приходите от ДДС отиват в държавата членка, в която живее потребителят. През септември Комисията започна *обществена консултация*, чрез която да се изберат начини за опростяване на плащанията по трансгранични сделки в сектора на електронната търговия в ЕС.

### Подобряване на трансграничната доставка на колетни пратки

Разходите за доставка на колетни пратки не са проблем само за потребителите, пазаруващи онлайн. Дружествата, продаващи продуктите си по електронен път, също се оплакват, че разходите за доставка са проблем. В рамките на стратегията за цифровия единен пазар Комисията насърчава *финансово достъпни висококачествени трансгранични доставки*, за да могат продуктите на предприятията да достигнат до потребителите по начин, който е от полза за всички страни по сделката. За да разбере от какво точно се нуждаят и какво искат гражданите и бизнесът, през май Комисията започна консултация по този въпрос.

## ЕЛЕКТРОННАТА ТЪРГОВИЯ: ОГРОМЕН ПОТЕНЦИАЛ, КОЙТО ОБАЧЕ ОСТАВА ДО ГОЛЯМА СТЕПЕН НЕИЗПОЛЗВАН НА РАВНИЩЕ ЕС



Източник: Joint Research Centre/Institute for Prospective Technological Studies, Digital economy working paper, 2015 г.

### Действия срещу блокирането на географски принципи

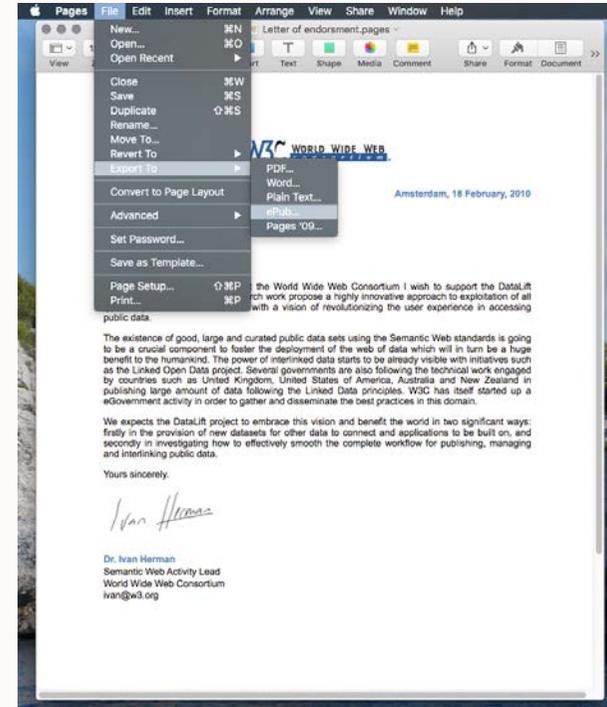
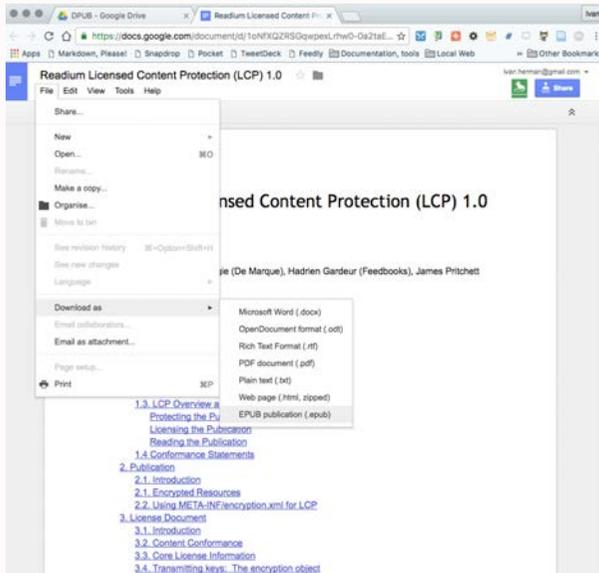
Много граждани, които пазаруват онлайн, стават жертва на блокиране на географски принципи — търговска практика, при която продавачите в интернет или

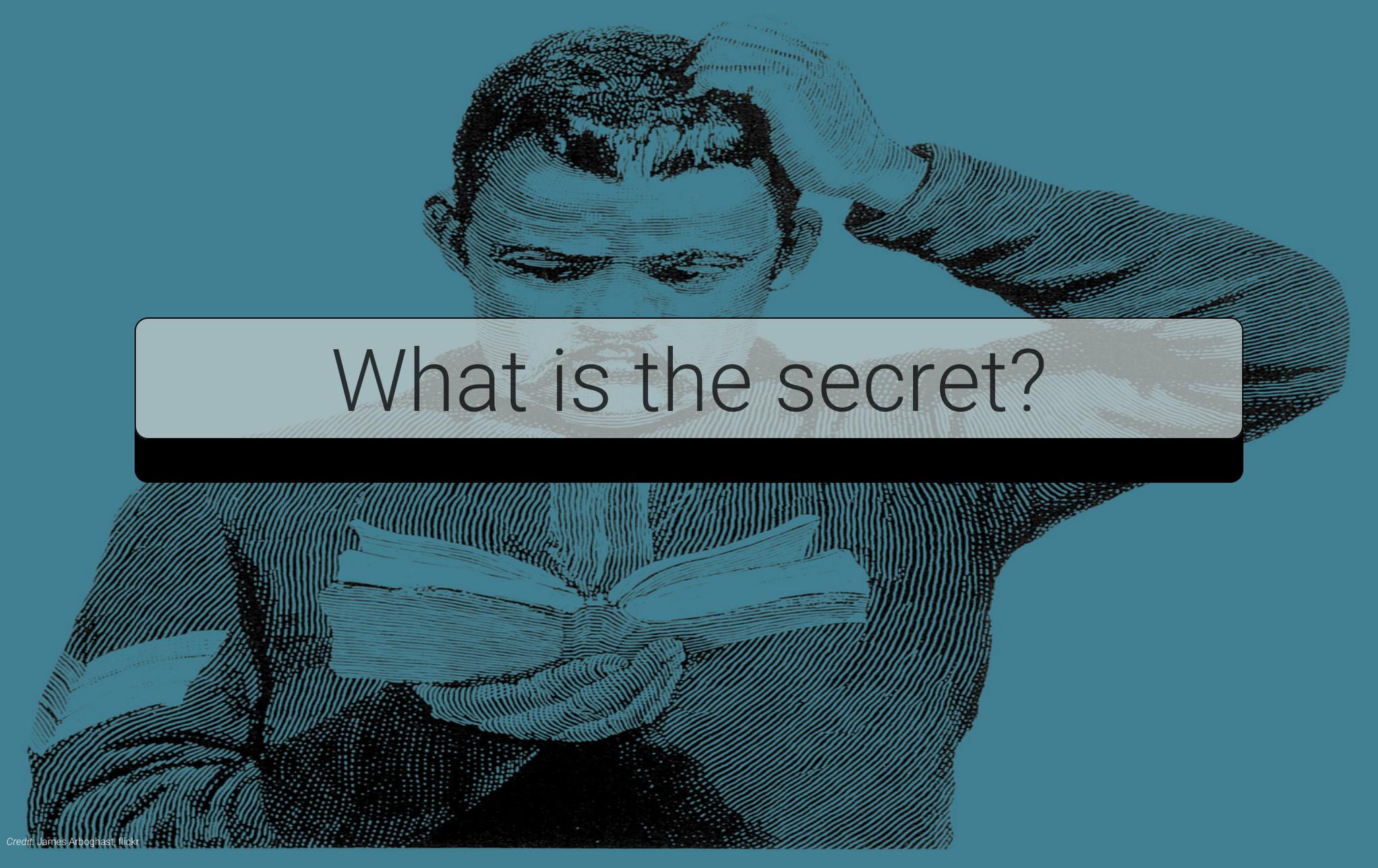
8 pages left in chapter

Максимално позициониране на потен...

Credit: European Commission – General Report 2015

# In fact, just about anything!



A black and white woodcut-style illustration of a man in a suit, looking down at an open book he is holding with both hands. His right hand is raised to his forehead, suggesting deep thought or a search for answers. The background is a solid dark color.

What is the secret?

## Well, at least one of the secrets...

---

- EPUB is based on Web Technologies like HTML, CSS, SVG, JS, etc.
  - collectively referred to as the “Open Web Platform” (OWP)
- I.e., no separate rendering, formatting, etc., specification had to be defined
- Instead, EPUB relies on specifications defined by W3C and deployed in browser engines

# Rough structure of an EPUB file

## Administrative files (manifest, metadata, rights,...)

```
<?xml version='1.0' encoding='utf-8'?>
<package
  xmlns="http://www.idpf.org/2007/opf"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  prefix="cc: http://creativecommons.org/ns#" unique-identifier="pub-id" version="3.0" xml:lang="en">
  <metadata>
    <dc:title id="title">
      PWP</dc:title>
    <dc:creator id="creator">
      Markus Gylling, Tzviya Siegman,
      Ivan Herman</dc:creator>
    <dc:language>en-US</dc:language>
    <rel="cc:license" />
  </metadata>
  <manifest>
    <item href="nav.xhtml" id="nav">
```

## Content (HTML, SVG, MathML,...)

```
<section>
  <h2>1. Our Vision</h2>

  <p>Our vision for EPUB-WEB is that portable documents become fully native citizens of the Open Web Platform. In this vision, the current format- and workflow-level separation between offline/portable (EPUB) and online (Web) document publishing is diminished to zero. These are merely two dynamic manifestations of the same publication: content authored with online use as the primary mode can easily be saved by the user for offline reading in portable document form. Content authored primarily for use as a portable document can be put online, without any need for re-factoring the content. Publishers can choose to utilize either or both of these publishing modes,
```

## Auxilliary files (PNG, CSS, JS,...)



```
div.issue, div.note, div.warning
{
  padding: 1em 1.2em 0.5em;
  margin: 1em 0;
  position: relative;
  clear: both;
}
span.note, span.issue,
span.warning
{ padding: .1em .5em .15em; }
.issue {
  border-color: #e05252;
```

Open Packaging Format (OPF)

# A good example

- The image on the right is displayed through that translucent “screen”
- This is not done through some separate image processing step
- Instead, it is based on some complex CSS features (implemented in browsers but also in EPUB readers)

CSS Secrets  
looks as follows, and its result can be seen in Figure 4.23:

```
body, main:before {  
  background: url("tiger.jpg") 0 / cover fixed;  
}  
  
main {  
  position: relative;  
  background: hsla(0,0%,100%,.3);  
  overflow: hidden;  
}  
  
main:before {  
  content: '';  
  position: absolute;  
  top: 0; right: 0; bottom: 0; left: 0;  
  filter: blur(20px);  
  margin: -30px;  
}
```

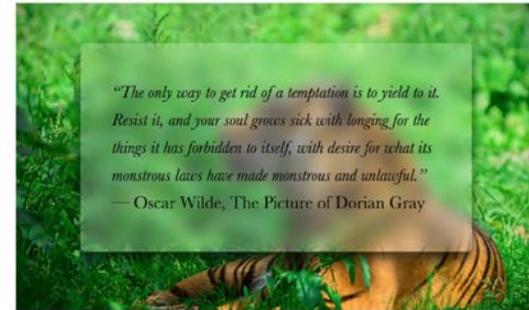


FIGURE 4.21 Blurring our pseudo-element almost works, but its less blurry on the edges, diminishing the frosted glass illusion

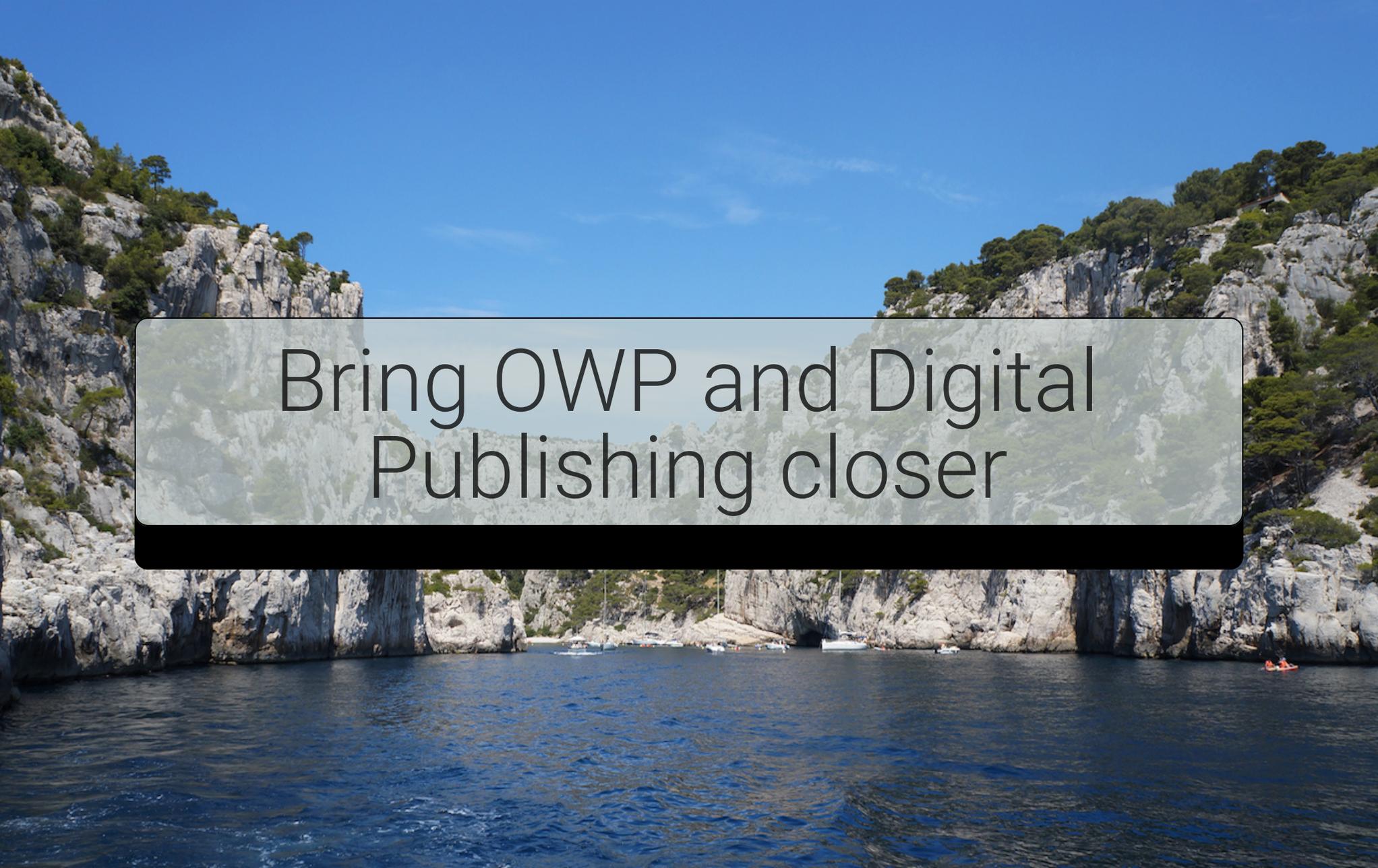
Bottom line: relationships of  
EPUB with OWP is  
fundamental

# Are we all done? I.e., are OWP and EPUB a perfect match?

---

There are two major areas that need work:

1. bring OWP and Digital Publishing closer
2. bring the Web and Digital Publishing closer

A scenic view of a rocky coastline with a blue sky and a body of water. A semi-transparent text box is overlaid in the center. The text box has a white background with rounded corners and a black border. The text inside is black and reads "Bring OWP and Digital Publishing closer".

Bring OWP and Digital  
Publishing closer

# Missing OWP features

---

- Some features, needed by publishing, are missing from OWP; examples:
  - pagination, page controls for headings, page numbers
  - graphics transition effects (needed in, e.g., mangas, cartoons, children books)
  - structuring information of complex documents
  - taking into account the typesetting requirements of different cultures (French, Russian, Arabic, Chinese, Japanese,...)
- W3C has a [Digital Publishing Interest Group](#) to look into these issues and cooperate with the relevant W3C groups

# Bring other OWP features to publishing

---

- For example:
  - dynamic tools and libraries, deployed on the Web, would be beneficial in publishing
  - use external services on the Web via extra protocols
  - new Web authentication or payment methods could change the way publications are accessed
  - new device API-s could be used by dedicated software on devices
  - etc.

Bringing these to publishing  
should (and will...) happen



Bring the Web and Digital  
Publishing closer

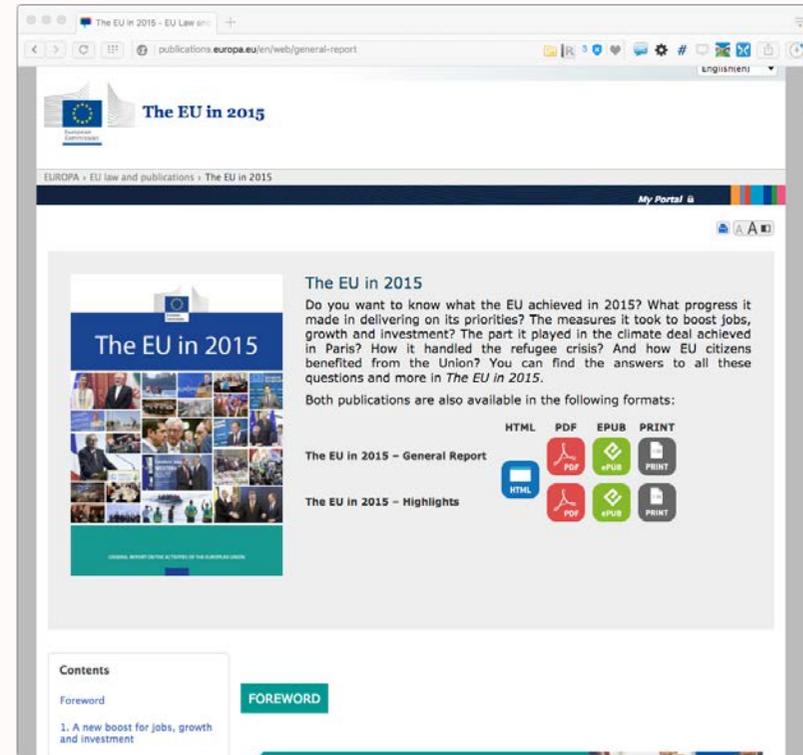
# What is, in fact, a (digital) book?

---

- It is a mean to transmit ideas to readers in various formats
- It is independent of format: the same “book” should be available in different format and media
  - printed book, digital books, Web page(s)...
- Today, a Digital Book is different than a Web page (conceptually and practically)

# What we get today...

- One can download an EPUB (and other formats) from a Web site...



The EU in 2015

Do you want to know what the EU achieved in 2015? What progress it made in delivering on its priorities? The measures it took to boost jobs, growth and investment? The part it played in the climate deal achieved in Paris? How it handled the refugee crisis? And how EU citizens benefited from the Union? You can find the answers to all these questions and more in *The EU in 2015*.

Both publications are also available in the following formats:

	HTML	PDF	EPUB	PRINT
The EU in 2015 – General Report				
The EU in 2015 – Highlights				

Contents

Foreword

1. A new boost for jobs, growth and investment

**FOREWORD**

Credit: Dump from EU publications' page

# What we get today...

- ...get a very traditional text to read...

## Foreword



Jean-Claude Juncker, President of the European Commission, gives his State of the Union address for 2015 in the European Parliament, Strasbourg, 9 September 2015.

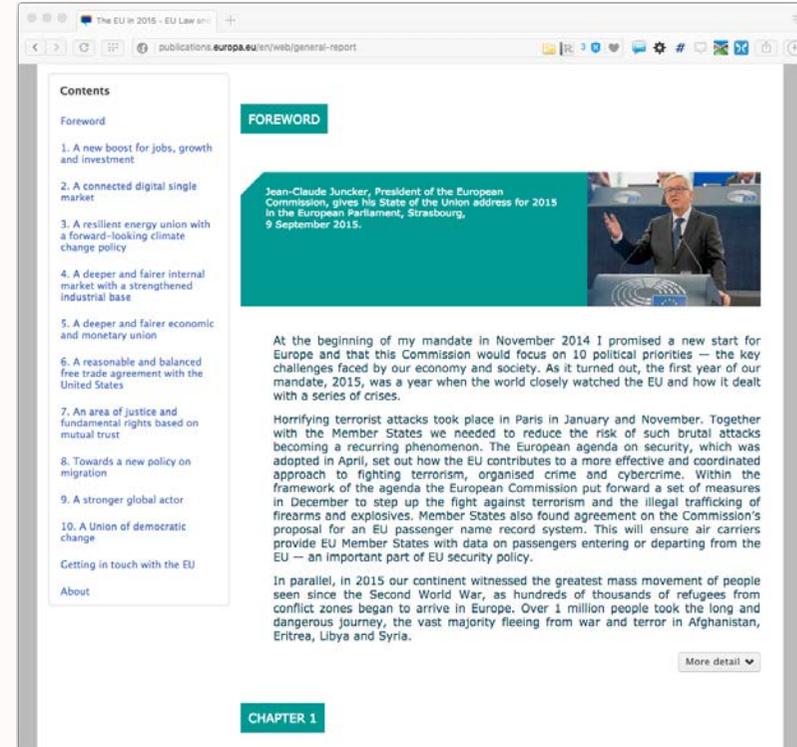
At the beginning of my mandate in November 2014 I promised a new start for Europe and that this Commission would focus on 10 political priorities — the key challenges faced by our economy and society. As it turned out, the first year of our mandate, 2015, was a year when the world closely watched the EU and how it dealt with a series of crises.

Horrifying terrorist attacks took place in Paris in January and November. Together with the Member States we needed to reduce the risk of such brutal attacks becoming a recurring phenomenon. The European agenda on security, which was adopted in April, set out how the EU contributes to a more effective and coordinated approach to fighting terrorism, organised crime and cybercrime. Within the framework of the agenda the European Commission put forward a set of measures in December to step up the fight against terrorism and the illegal trafficking of firearms and explosives. Member States also found agreement on the Commission's proposal for an EU passenger name record system. This will ensure air carriers provide EU Member States with data on passengers entering or departing from the EU — an important part of EU security policy.

In parallel, in 2015 our continent witnessed the greatest mass movement of people seen since the Second World War, as hundreds of thousands of refugees from conflict

# What we get today...

- ...but one can access a different (and possibly nicer) Web page online...
- *This should not be the case!*



The screenshot shows a web browser displaying the 'Foreword' section of the 'The EU in 2015' report. The page is titled 'FOREWORD' and features a green header with a photo of Jean-Claude Juncker, President of the European Commission, giving his State of the Union address. The main text discusses the challenges faced by the EU in 2015, including terrorist attacks in Paris and the refugee crisis. A 'More detail' button is visible at the bottom right of the text area. A 'CHAPTER 1' label is at the bottom of the page.

Contents

Foreword

1. A new boost for jobs, growth and investment
2. A connected digital single market
3. A resilient energy union with a forward-looking climate change policy
4. A deeper and fairer internal market with a strengthened industrial base
5. A deeper and fairer economic and monetary union
6. A reasonable and balanced free trade agreement with the United States
7. An area of justice and fundamental rights based on mutual trust
8. Towards a new policy on migration
9. A stronger global actor
10. A Union of democratic change

Getting in touch with the EU

About

**FOREWORD**

Jean-Claude Juncker, President of the European Commission, gives his State of the Union address for 2015 in the European Parliament, Strasbourg, 9 September 2015.

At the beginning of my mandate in November 2014 I promised a new start for Europe and that this Commission would focus on 10 political priorities – the key challenges faced by our economy and society. As it turned out, the first year of our mandate, 2015, was a year when the world closely watched the EU and how it dealt with a series of crises.

Horrifying terrorist attacks took place in Paris in January and November. Together with the Member States we needed to reduce the risk of such brutal attacks becoming a recurring phenomenon. The European agenda on security, which was adopted in April, set out how the EU contributes to a more effective and coordinated approach to fighting terrorism, organised crime and cybercrime. Within the framework of the agenda the European Commission put forward a set of measures in December to step up the fight against terrorism and the illegal trafficking of firearms and explosives. Member States also found agreement on the Commission's proposal for an EU passenger name record system. This will ensure air carriers provide EU Member States with data on passengers entering or departing from the EU – an important part of EU security policy.

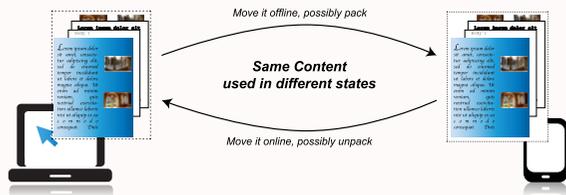
In parallel, in 2015 our continent witnessed the greatest mass movement of people seen since the Second World War, as hundreds of thousands of refugees from conflict zones began to arrive in Europe. Over 1 million people took the long and dangerous journey, the vast majority fleeing from war and terror in Afghanistan, Eritrea, Libya and Syria.

[More detail](#)

**CHAPTER 1**

Credit: Dump from EU publications' page

# *“This should not be the case!”* what does this mean?



*Separation between publishing as Web sites and in an offline package should be diminished to zero*

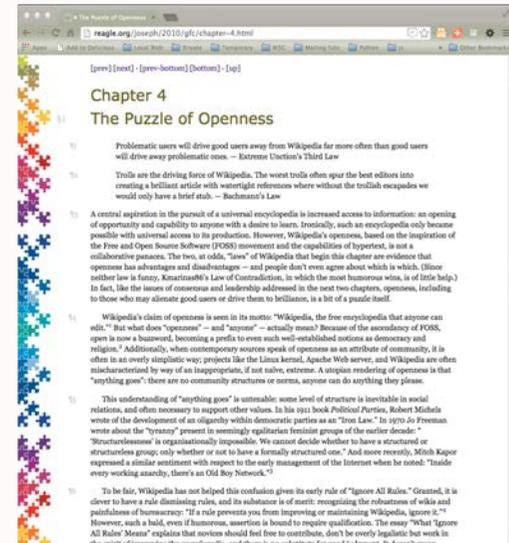
- This means:
  - it should be possible to load the publication content into a browser or a specialized reader, whatever the user prefers
  - it should be possible to read the book either offline or online, whatever the circumstances dictate
  - contents could be authored regardless of where they are used

Why ?



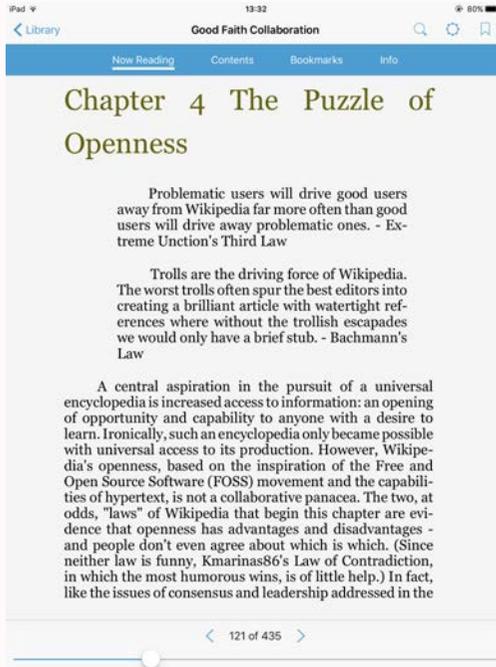
# For example: book in a browser

- On a desktop I may want to read a book just like a Web page:
  - easily follow a link “out” of the book
  - create bookmarks to “within” a page in a book
  - use useful plugins and tools that my browser may have
  - create annotations
  - sometimes I may need the computing power of my desk-top for, e.g., powerful text analysis



Credit: Joseph Reagle: "Good Faith Collaboration", PhD Thesis, MIT Press

# For example: book in a browser



Credit: Joseph Reagle: "Good Faith Collaboration", PhD Thesis, MIT Press

- But, at other times, I may also want to use a small dedicated reader device to read the book on the beach...
- All these on *the same* book (not conversions from one format to the other)!

## For example: I may not be online...

---

- I may find an article on the Web that I want to review, annotate, etc., while commuting home on a train
- I want the results of the annotations to be back online, when I am back on the Internet
- note: browsers may have an “archiving” possibility, but they are not interoperable



*Credit: Bryan Ong, Flickr*

# For example: scholarly publishing

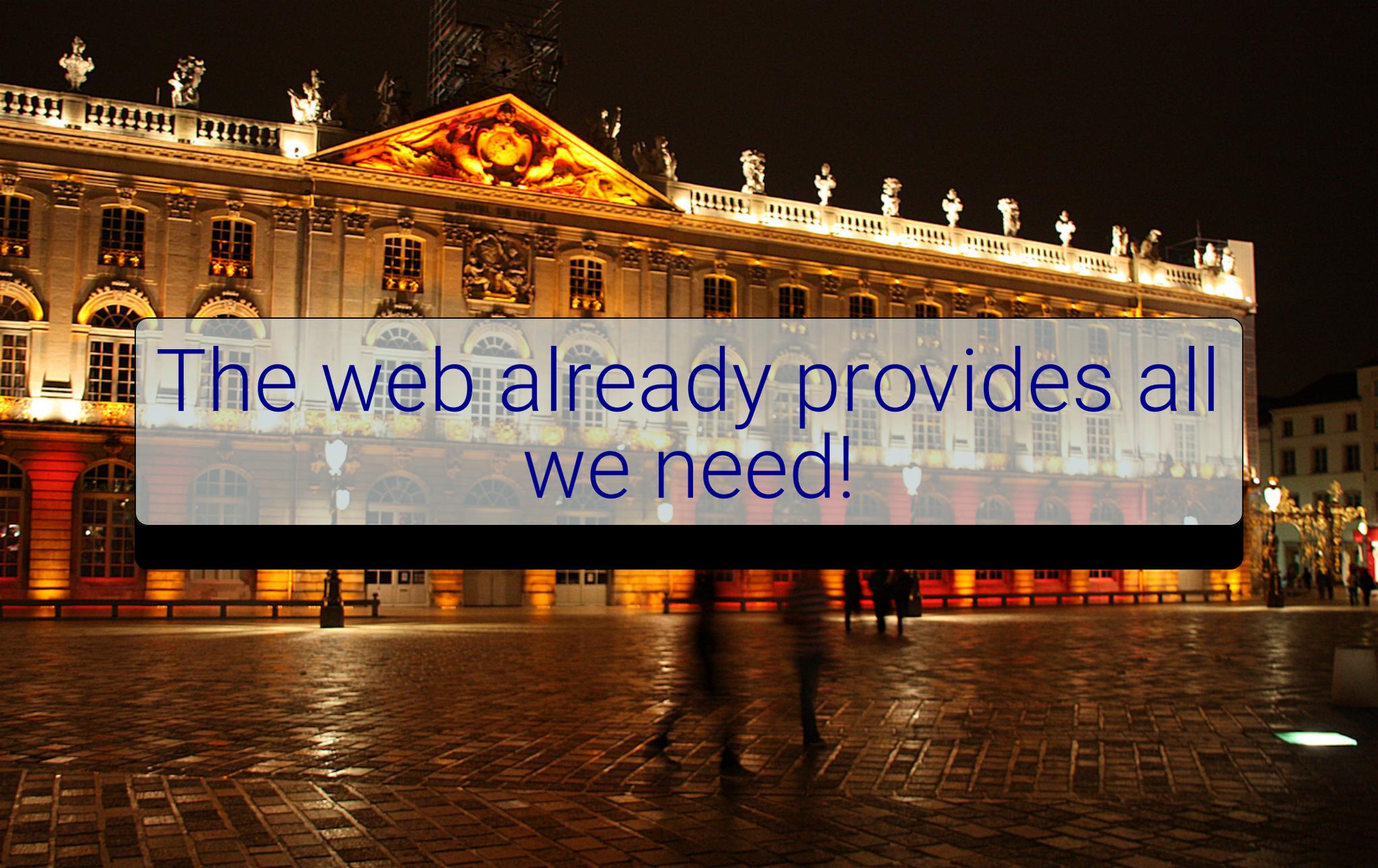


Credit: Jullien Colomb et. al: "Sub-straains of *Dr osophila Canton-S...*", F1000Research

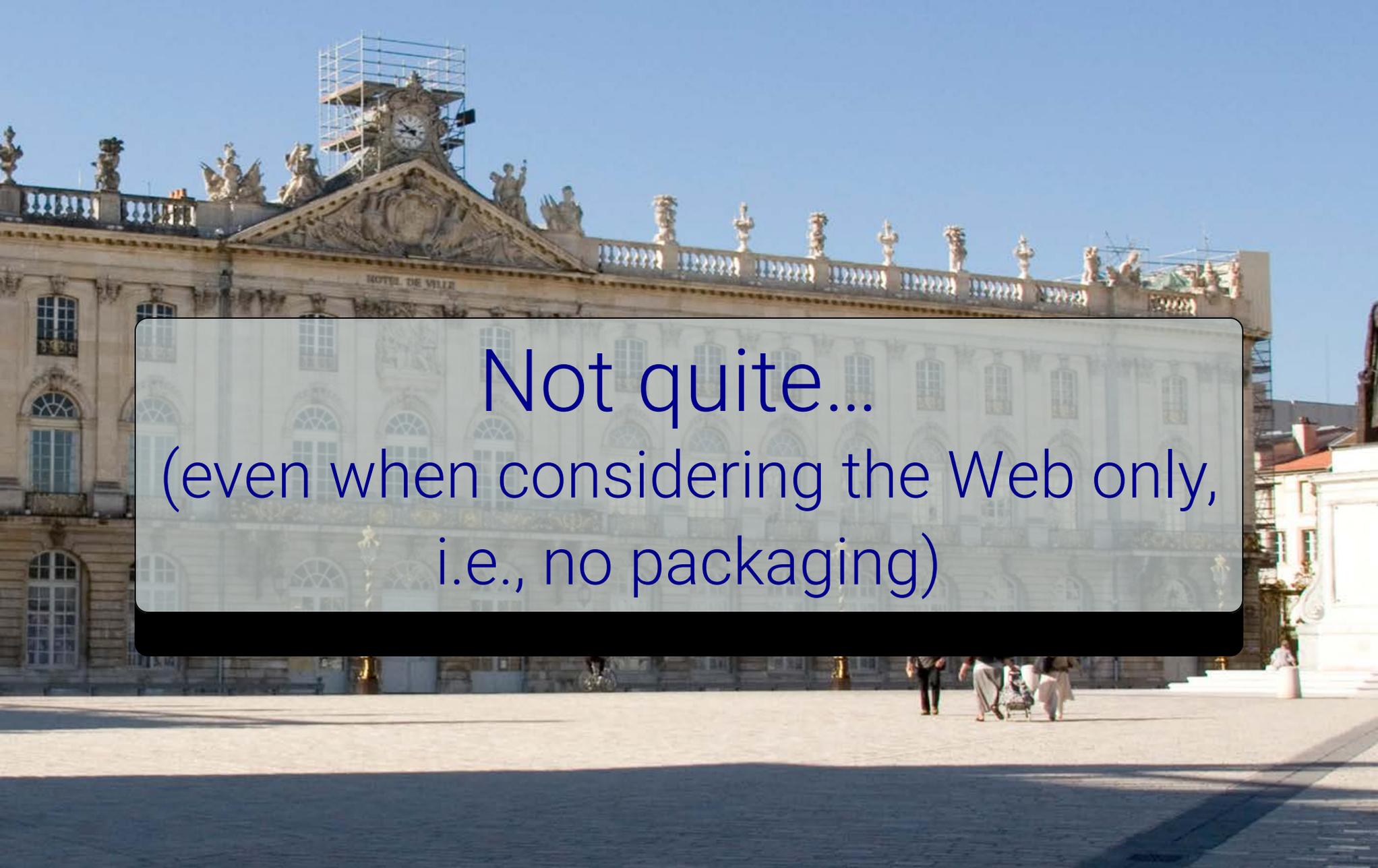
- Paper may be published on-line, but people may want to download it for offline use
- The format of the paper should be adaptable to reading environment
  - no two column, fixed layout file that cannot be handled on an my tablet...
- The “paper” may also contain video, audio, data, programs...
  - scholarly publishing is not text only any more!



But... why not simply rely  
*only* on the current Web?  
(with some facilities for offline)



The web already provides all we need!



Not quite...

(even when considering the Web only,  
i.e., no packaging)

# Need for the concept of a “publication” of many resources

---

- A publication may consist of many resources (text, images, videos, etc)
- On the Web, these are linked together, but they keep their complete “independence”
- For publishing the abstract “work” should be a single, conceptual unit on the Web: a **Web Publication (WP)**

# Why do we need the WP concept?

---

- The whole book should be archived in a national library
- The publication has to be checked for integrity as a whole (e.g., in a legal publication)
- Sharing a publication with a friend requires the concept of the publication as a whole
- Publishing and library business models are based on the concept of “works”

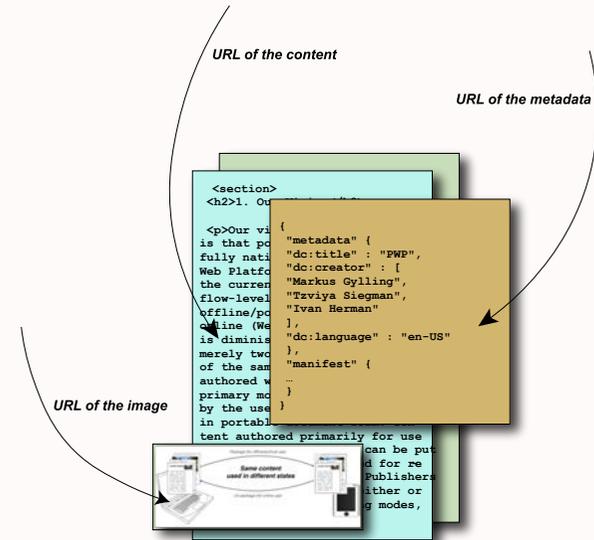
# Why do we need the WP concept?

---

- Value counters (counters, section numbering, footnotes, endnotes, etc.) should be consistent across the publication
- Any search, text analysis, etc., facility for a book should have access to the full publication
- User preferences should apply to the “work” rather than to a single Web page

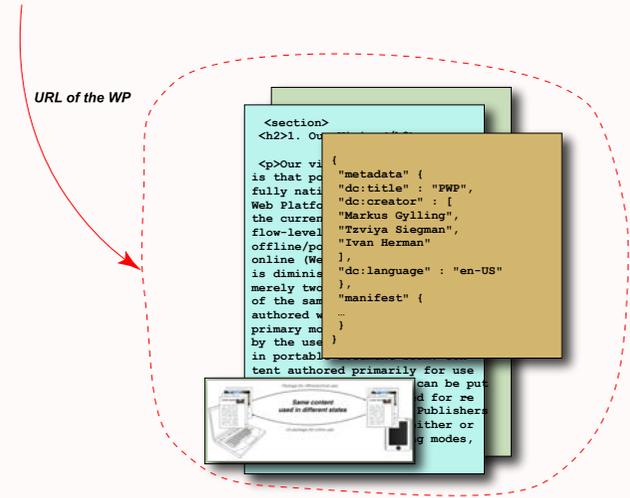
# How does that translate to the Web?

- The current Web builds on the notion of a single resource:
  - HTML source, metadata, CSS style sheet, etc.
  - each has its own URL
- Rendering is based on the interoperation of many such resources



# How does that translate to the Web?

- A WP is a collection of Web resources, considered as a single Web resource
  - this means there is a unique Web address for the *collection*
  - the same WP can be offline or online with the same address
- Web Publications are part of the Web!



# An additional concept: a “WP Processor”

---

- A conceptual, client-side processor (add-on, extension, etc.) that performs WP specific operations
  - search, unified counters and section numbers, displaying table of contents, publication-specific interactions,...
- It “hides” the WP online/offline state
  - the main rendering part operates *as if* it was connected to the Web

# An internal representation may also be needed

---

- At the moment, the internal representation of (HTML) resources is based on the concept of a “Document” in the “Document Object Model (DOM)”
  - this underpins all programmatic access to content on the Web
- There *may* be a need for a programmatic interface on top of Documents in the DOM, representing a *collection* of documents

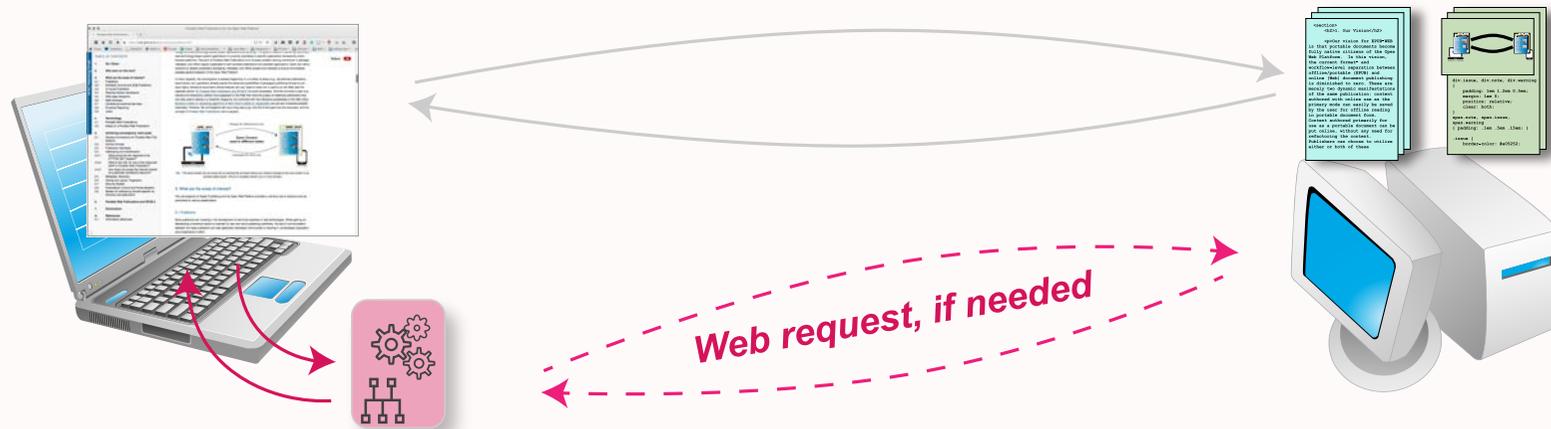


Architectural challenge:  
handling online/offline

# Envisioned architecture: online



# Envisioned architecture: offline



**Requests handled by the WP Processor  
content possibly cached**

Is this approach at all  
feasible?

# Advances in modern browsers: Web and Service Workers

---

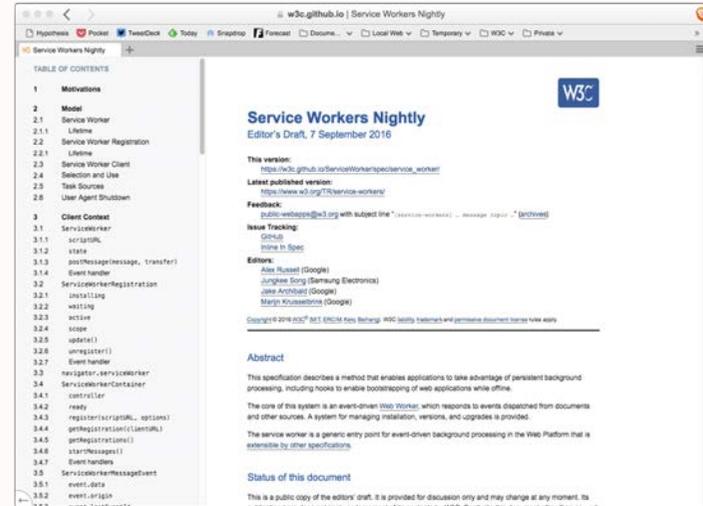
- Web Worker: a truly parallel thread within the browser
- A Service Worker is a special type of Web Worker, with additional features:
  - it is a *programmable network proxy*: the renderer's network calls are caught and the request/answer can be modified on-the-fly behind the scenes
  - it has an interface to handle a local cache for networked data
  - it will stay alive even if the user moves away from the main page, and can be accessed later if he/she returns to it

Work in progress

A WP Processor can be  
implemented using Service  
Workers

# Service Workers are coming...

- Browser engines are heading towards Service Workers today
- a major technology to reconcile Mobile Applications and Web Developments
- I.e, Web Publications may build on an evolution that is already happening



# An example of online/offline book with Service Workers

---



# Manifests

---

- A WP needs a number of extra information
  - different type of metadata (author, title, publisher, etc.)
  - list of resources *within* a WP
  - table of content information, i.e., default reading order
  - unique identifier (ISBN, DOI, etc.)
  - references to more complex metadata (e.g., ONIX data)
- The traditional term is “manifest” for such information
- A detailed manifest must be specified for WP-s



# Packaged Web Publications

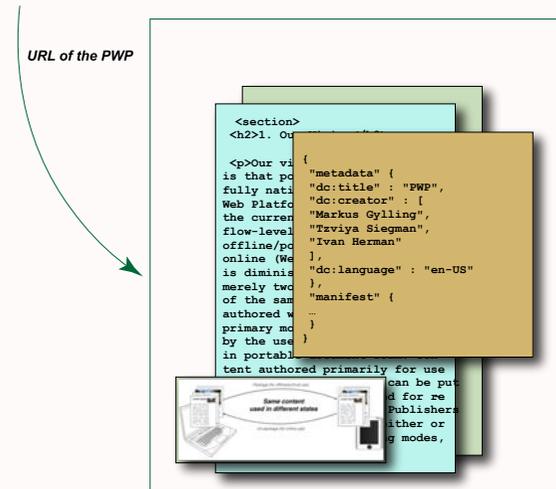
# Packaged Web Publications (PWP)

---

- For a large number of use cases Web Publications (essentially, online/offline reading) is all they really need
  - e.g., scholarly journals, online magazines, brochures
- But a number of use cases rely on an extra step: to “pack” the content of a WP into a single file
  - archival, library distributions,...
  - extension of the current business model of trade publishers (built on EPUB)

# A layer “on top” of WP-s

- Packaging means collecting all the content of a WP into a single file
- Some additional issues to be handled:
  - special entries to the manifest
  - archival format(s)
  - etc.
- A specific “profile” for WP may have to be defined for a PWP



# Structure of an EPUB3 file

## Administrative files (manifest, metadata, rights,...)

```
<?xml version='1.0' encoding='utf-8'?>
<package
  xmlns="http://www.idpf.org/2007/opf"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  prefix:cc="http://creativecommons.org/ns#" unique-identifier="pub-id" version="3.0" xml:lang="en">
  <metadata>
    <dc:title id="title">
      PWP</dc:title>
    <dc:creator id="creator">
      Markus Gylling, Tzviya Siegman,
      Ivan Herman</dc:creator>
    <dc:language>en-US</dc:language>
    <rel="cc:license" />
  </metadata>
  <manifest>
    <item href="nav.xhtml" id="nav">
```

## Content (HTML, SVG, MathML,...)

```
<section>
  <h2>1. Our Vision</h2>

  <p>Our vision for EPUB-WEB is that portable documents become fully native citizens of the Open Web Platform. In this vision, the current format- and workflow-level separation between offline/portable (EPUB) and online (Web) document publishing is diminished to zero. These are merely two dynamic manifestations of the same publication: content authored with online use as the primary mode can easily be saved by the user for offline reading in portable document form. Content authored primarily for use as a portable document can be put online, without any need for re-factoring the content. Publishers can choose to utilize either or both of these publishing modes,
```

## Auxilliary files (PNG, CSS, JS,...)



```
div.issue, div.note, div.warning
{
  padding: 1em 1.2em 0.5em;
  margin: 1em 0;
  position: relative;
  clear: both;
}
span.note, span.issue,
span.warning
{ padding: .1em .5em .15em; }

.issue {
  border-color: #e05252;
```

Open Packaging Format (OPF)

# A Packaging of a Web Publication

---

- A Packaged Web Publication may have a very similar structure
- But moving to the Web would probably require to rethink the format of the administrative files
  - e.g., manifest would be in JSON
  - structure of metadata may differ
  - maybe re-think the packaging format itself
  - etc.

# PWP packaging structure

Administrative files  
(manifest, metadata, rights,...)

```
{  
  "metadata" {  
    "dc:title" : "PWP",  
    "dc:creator" : [  
      "Markus Gylling",  
      "Tzviya Siegman",  
      "Ivan Herzan:"  
    ],  
    "dc:language" : "en-US"  
  },  
  "manifest" {  
    ...  
  }  
}
```

Content  
(HTML, SVG, MathML,...)

```
<section>  
  <h2>1. Our Vision</h2>  
  
  <p>Our vision for EPUB-WEB  
is that portable documents become  
fully native citizens of the Open  
Web Platform. In this vision,  
the current format- and work-  
flow-level separation between  
offline/portable (EPUB) and  
online (Web) document publishing  
is diminished to zero. These are  
merely two dynamic manifestations  
of the same publication: content  
authored with online use as the  
primary mode can easily be saved  
by the user for offline reading  
in portable document form. Content  
authored primarily for use  
as a portable document can be put  
online, without any need for re-  
factoring the content. Publishers  
can choose to utilize either or  
both of these publishing modes,
```

Auxilliary files  
(PNG, CSS, JS,...)



The diagram shows a central text box with the text "Same content used in different states" and arrows pointing to three devices: a laptop, a tablet, and a smartphone. Above the devices, it says "Package for offline/online use".

```
div.issue, div.note, div.warning  
{  
  padding: 1em 1.2em 0.5em;  
  margin: 1em 0;  
  position: relative;  
  clear: both;  
}  
span.note, span.issue,  
span.warning  
{ padding: .1em .5em .15em; }  
  
.issue {  
  border-color: #e05252;
```

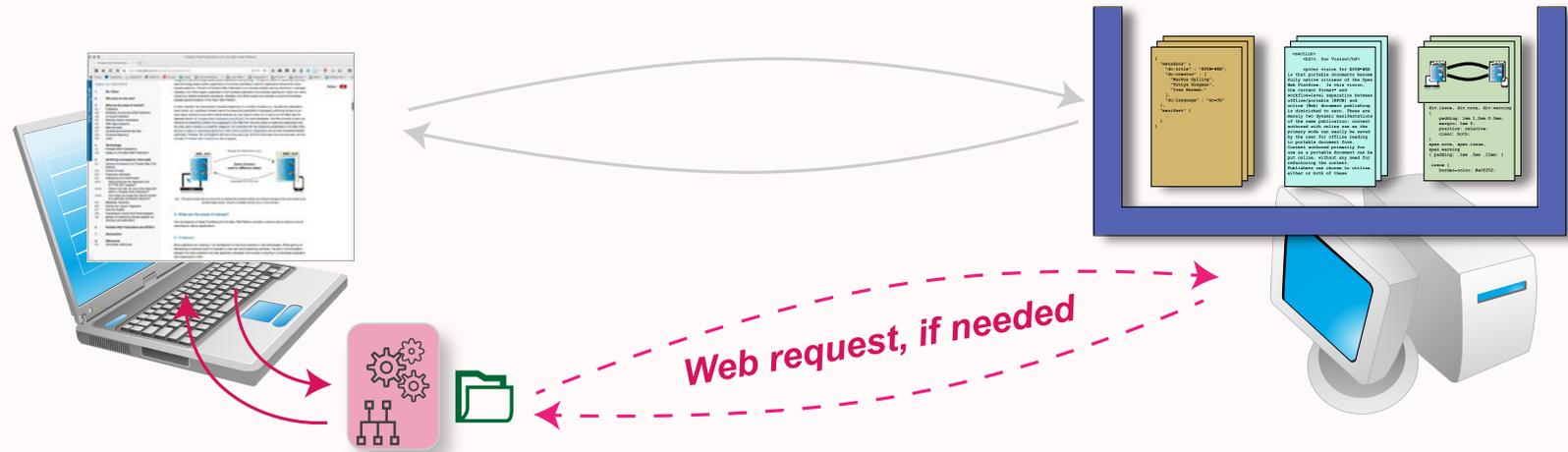
Web Packaging Format

# A PWP Processor

---

- PWP-s may be handled by dedicated applications
- But the concept of a WP processor can be extended to a PWP processor, unpacking content on the fly
  - security considerations must be handled, though

# A PWP Processor



**Requests handled by the PWP Processor  
content unpacked on-the-fly  
and possibly cached**

Technical challenge:  
addressing, identification



# Is it “addressing” or “identification”?

---

- These two “roles” are different
- The usual situation:
  - some form of a URI is used to (*uniquely*) *identify* a resource
  - an HTTP(S) URL is used to address (or “locate”) a resource on the Web
- In many cases the two roles coincide, but not always
  - e.g., for a digital Book :
    - [URN:ISBN:1-56592-521-1](#) identifies the publication
    - <http://www.ex.org/ex.wp> addresses a particular copy

## Is it “addressing” or “identification”?

---

- Identification is handled by a number of other organizations (like the [DOI foundation](#), the [International ISBN Agency](#), etc.)
- The work on (P)WP has to concentrate on locators (i.e., addressing)

# What does a Web request return for L?

---

- Possibilities are:
  - the package that *includes* the manifest at some predefined place (in case this is a Packaged Web Publication)
  - an HTML file with a link to a manifest (through a `<link>` element)
  - an HTML file with an *embedded* manifest (through a `<script>` element)
  - an HTML file with a link to a manifest in the `Link` header of the `HTTP` response
- Unclear whether all of these are necessary and/or feasible



# Ergonomics differences

Gingembre  
50 gr → 4 €  
100 gr → 7 €

Paprika  
50 gr → 4 €  
100 gr → 7 €

Curry de Madras  
50 gr → 4 €  
100 gr → 7 €

Spécial Pomme de Terre  
25 gr → 4,50 €  
50 gr → 8 €  
100 gr → 15 €

Raz El Anouth Rouge  
25 gr → 2 €

Chili  
50 gr → 4 €  
100 gr → 7 €

Piment de Cayenne

Café  
25 gr → 2 €  
50 gr → 3,50 €

Special Pizza  
25 gr → 2 €  
50 gr → 3,50 €

Paprika  
25 gr → 2 €  
50 gr → 3,50 €

Basilic  
25 gr → 2 €  
50 gr → 3,50 €

Origan  
25 gr → 2 €  
50 gr → 3,50 €

Thym  
25 gr → 2 €  
50 gr → 3,50 €

Romarin  
25 gr → 2 €  
50 gr → 3,50 €

Margolaine  
25 gr → 2 €  
50 gr → 3,50 €

Margolaine  
25 gr → 2 €  
50 gr → 3,50 €

The best  
50 gr → 4 €  
100 gr → 7 €

The best  
50 gr → 4 €  
100 gr → 7 €

The best  
50 gr → 4 €  
100 gr → 7 €

The best  
50 gr → 4 €  
100 gr → 7 €

The best  
50 gr → 4 €  
100 gr → 7 €

The best  
50 gr → 4 €  
100 gr → 7 €

The best  
50 gr → 4 €  
100 gr → 7 €

The best  
50 gr → 4 €  
100 gr → 7 €

Margolaine  
25 gr → 2 €  
50 gr → 3,50 €

Piment de Cayenne

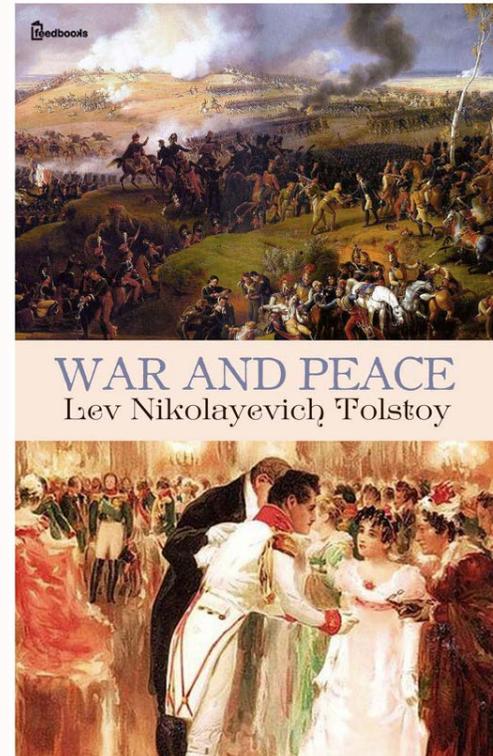
Piment de Cayenne

Piment de Cayenne

# Book reading needs a different approach to ergonomics

---

- Typical example: pagination
  - for short text, pagination is not really important
  - but what about reading, say, the War and Peace?
  - currently not available or used on the Web
- Pagination isn't only books: slides, cards, and other "boxes" use the same functionality

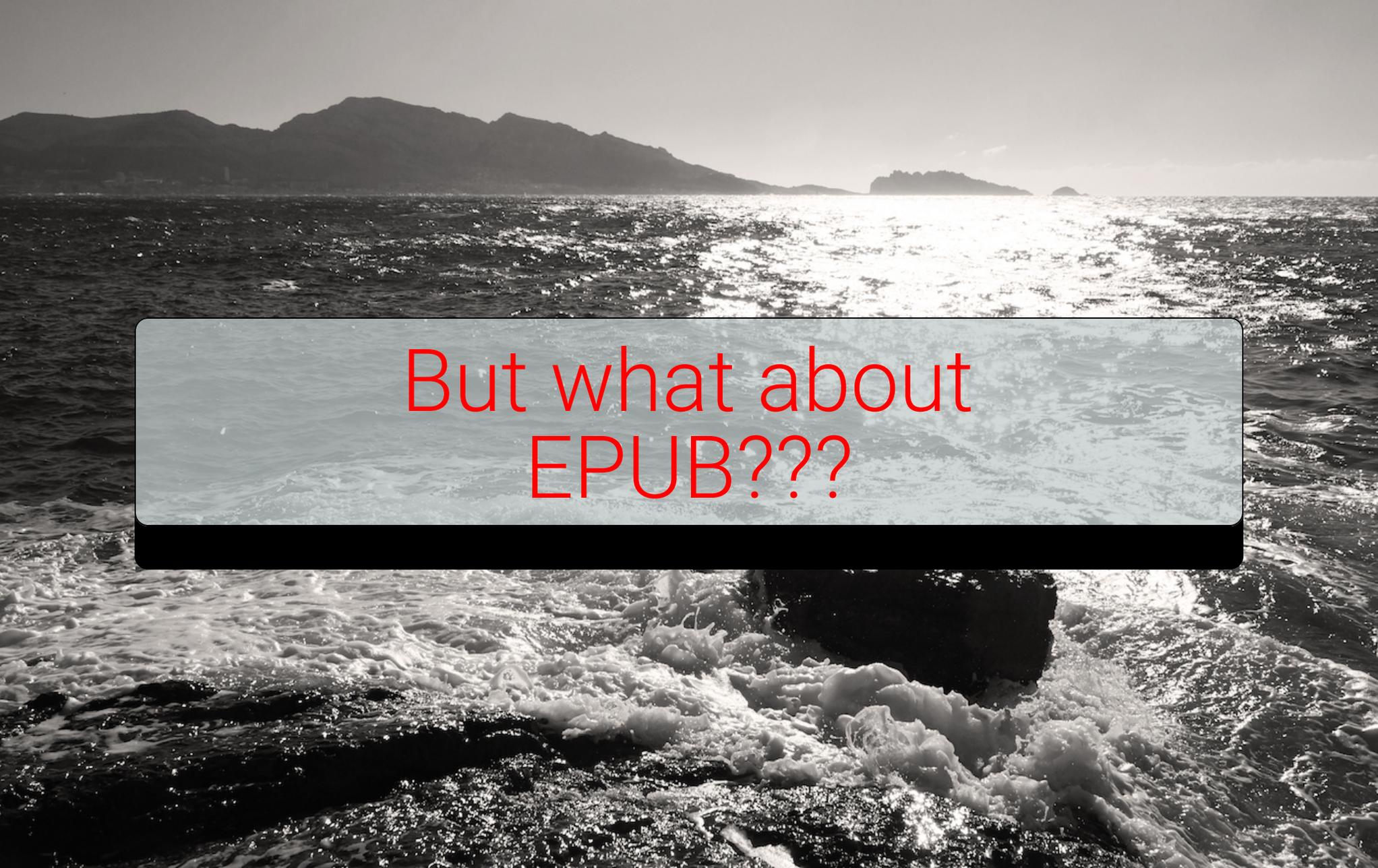


Credit: Lev Tolstoy: "War and Peace", feedbooks

# Personalization

---

- What is the level of user control of the presentation?
- The Web and eBook traditions are vastly different:
  - in a browser, the Web designer is in full control
    - CSS alternate style sheets or user style sheets are hardly in use
    - some user interface aspects can be controlled but only for the browser as a whole
  - in an eBook reader, there is more user control
    - foreground/background color
    - choice of fonts
- There is a need to reconcile these traditions



But what about  
EPUB???

# PWP vs. EPUB3.1

---

- PWP does *not* replace EPUB 3
- It is, instead, part of a future evolution, a “vision” for the years to come

# Most things are the same!

---

- Important to note that the content of a PWP and EPUB 3 are essentially identical
  - usage of HTML, CSS, SVG, etc.
- The main change is the “administrative” aspect of the publication (locators, manifest, etc)

# Development process

---

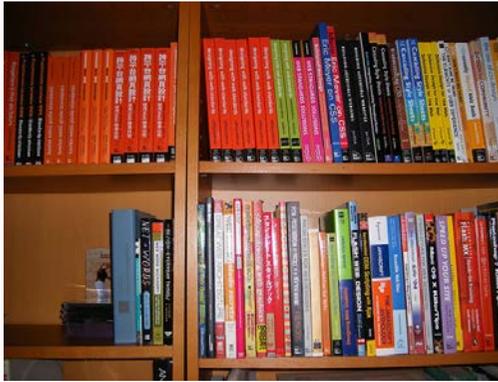
- EPUB was developed at the IDPF, Web Technologies were developed at the W3C
- Strong cooperation exists since 2013 (e.g., in W3C's Digital Publishing Interest Group)
  - many of the ideas around PWP were developed at W3C, with strong presence of experts from IDPF
- The plan is that IDPF and W3C would merge starting 2017
  - maintenance of EPUB 3.1 will continue at W3C
  - future development (of PWP and others) will be part of one or more W3C Working Groups

The image is a composite of two photographs. The top half shows a clear blue sky with two modern skyscrapers. The one on the left is a tall, rectangular building with a grid-like facade. The one on the right is a shorter, cylindrical building with a glass and metal facade. The bottom half shows a street scene with traditional brick row houses, trees with autumn foliage, and several cars parked along the street. A semi-transparent blue banner with a black border is overlaid across the middle of the image, containing the text "Other synergy effects of convergence" in a dark blue, serif font.

# Other synergy effects of convergence

# Advantage for the publishers' community

---

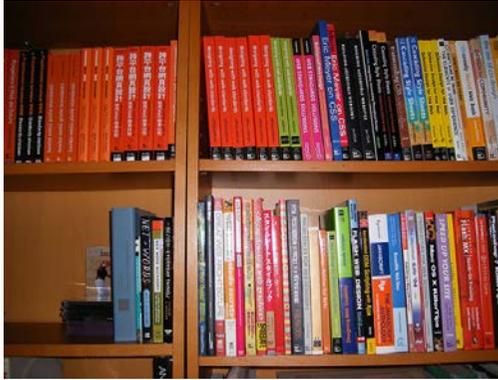


Credit: Jeffrey Zeldman, Flickr

- The main interest of publishers is to produce, edit, curate, etc., content
- Publishers would have an easier access to a large audience
- Publishers have invested heavily into technology developments, but the Web developers' community can complement that with a wider reach and perspective

# Advantage for the publishers' community

---

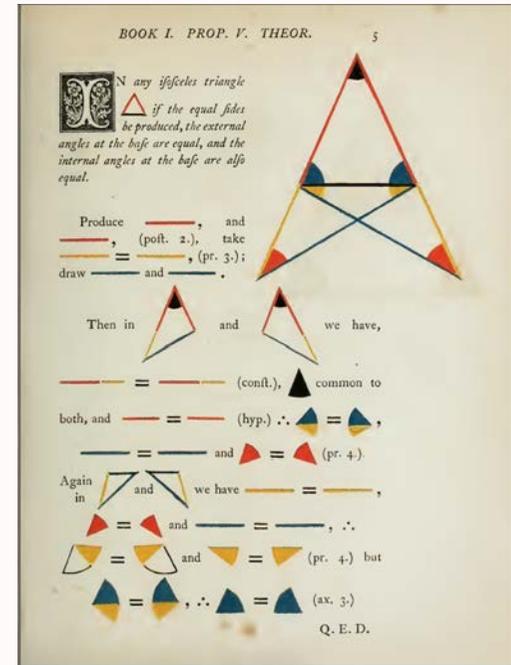


Credit: Jeffrey Zeldman, Flickr

- Working closely with Web developers avoids re-inventing wheels
  - e.g., API-s to access system resources or utilities, like index database, battery status API, real-time communication, geolocation,...
- Aligning more on OWP means that specialized readers or Web sites, software, services, etc, can rely on it

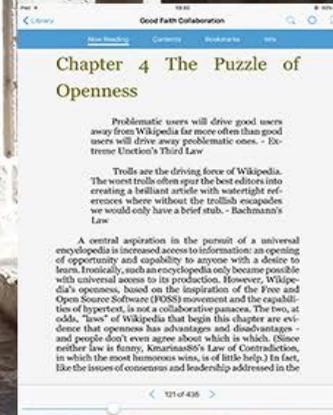
# Advantage for the Web community

- Publishers have experience in:
  - ergonomics, typography, aesthetics...
  - publishing long texts, with the right readability and structure
- Workflow for producing complex content
- In some sense, a new type of content comes to the fore



Credit: Oliver Byrne's edition of Euclid, University of British Columbia

# To conclude: Let us create real publications on the Web!



## Some references

---

Latest PWP Use Cases and Requirements draft:

<https://w3c.github.io/dpub-pwp-ucr/>

Latest PWP Editors' draft:

<https://w3c.github.io/dpub-pwp/>

PWP Issue list:

<https://github.com/w3c/dpub-pwp/issues>

constantly evolving...

One more thing...

This is *not* how Web development works at W3C...





...it is more like that!





I.e., join W3C to help things moving forward!

# Thank you for your attention!

---

This presentation:

<https://w3c.github.io/dpub/2016/SWIB/>

(PDF is also available for download)

My contact:

ivan@w3.org