

Policy Aware Content Reuse on the Web

Oshani Seneviratne

PhD Student at DIG, MIT CSAIL

Cambridge Semantic Web Gathering
September 8, 2009

Content Reuse on the Web

- There's so much content on the Web
 - 3.6 billion images on **flickr**
 - 20 hours of video uploaded every minute on **You Tube**
Broadcast Yourself™
- Content reuse is important
 - Prevents redundant work
 - Promotes creativity
- Several different types of policy mechanisms for content reuse
 - Upfront enforcement such as DRM
 - Rights expression such as **cc creative commons**

Creative Commons Licenses

- Can be expressed in human readable and machine readable formats
 - CC supports very user friendly icons and license deed pages
 - CC licenses can be expressed in RDF (ccREL spec)
- Can be deployed on a range of media
 - CC licenses can be applied to images, audio, video and text
- Large community
 - For e.g. there are 100 million CC licensed images on **flickr**
 - Most search engines support finding CC licensed content


Giving your content a Creative Commons License


- Most sites have a CC license option:
- CC offers a license chooser hosted at:
<http://creativecommons.org/choose>




- Generates a snippet of XHTML with RDFa
- Includes *cc:AttributionName* & *cc:AttributionURL*
- Can extend using *cc:morePermissions*


Allow commercial uses of your work?


☒ Yes 


☐ No 


Allow modifications of your work?

☒ Yes 

☐ Yes, as long as others share alike 

☐ No 

Jurisdiction of your license 

United States 

Anatomy of a CC License

Human readable form

A light gray rounded rectangle with a black border and a subtle drop shadow, containing the word "content".

content



This work by [John Doe](#) is licensed under a [Creative Commons Attribution 3.0 United States License](#).
Based on a work at <http://www.example-source.com>.
Permissions beyond the scope of this license may be available at <http://morepermissions.com>

Anatomy of a CC License

Under the hood (HTML + RDFa)

```

▼ <div xmlns="http://www.w3.org/1999/xhtml" xmlns:dc="http://purl.org/dc/terms/" xmlns:cc="http://creativecommons.org/ns#" about="something">
  ▼ <a rel="license" href="http://creativecommons.org/licenses/by/3.0/us/">
    
  </a>
  <br>
  <span href="http://purl.org/dc/dcmitype/Text" property="dc:title" rel="dc:type">This work</span>
  " by "
  <a href="http://www.example.com" property="cc:attributionName" rel="cc:attributionURL">John Doe</a>
  " is licensed under a "
  <a rel="license" href="http://creativecommons.org/licenses/by/3.0/us/"> Creative Commons Attribution 3.0 United States License </a>
  " "
  <br>
  " Based on a work at "
  <a href="http://www.examples-source.com" rel="dc:source">http://www.examples-source.com</a>
  " "
  <br>
  " Permissions beyond the scope of this license may be available at "
  <a href="http://more-permissions.com" rel="cc:morePermissions">http://morepermissions.com</a>
</div>

```

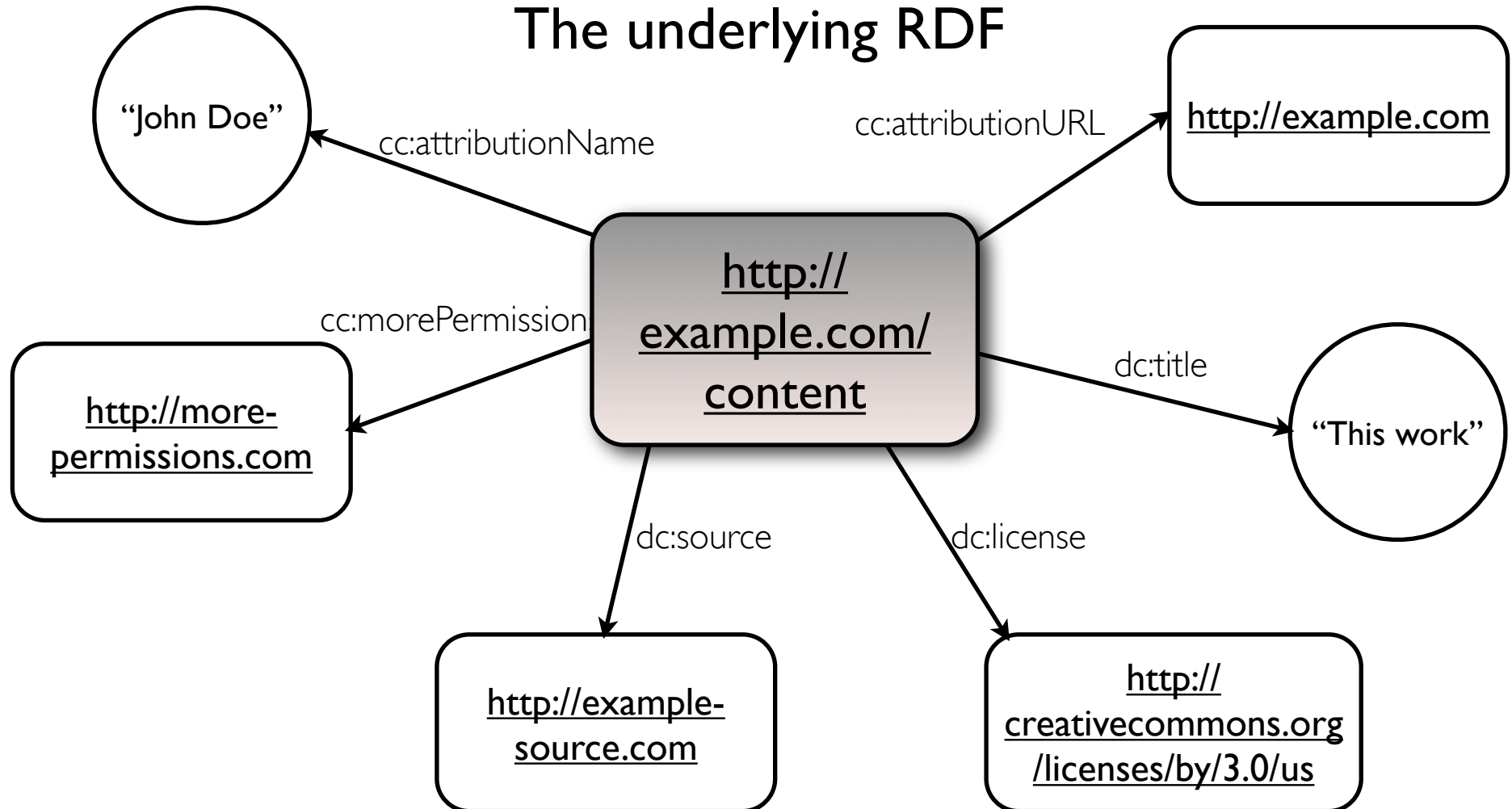
Legend

Attribute

Value

Anatomy of a CC License

The underlying RDF



CC Deeds using RDFa and Javascript



Attribution 3.0 United States

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Permissions beyond the scope of this public license are available at [more-permissions.com](http://creativecommons.org/licenses/by/3.0/us/permissions)



Permissions

Under the following conditions:



Attribution — You must attribute this work to [John Doe](#) (with link).

Attribute this work:

`<div xmlns:cc="http://creativecommons.org/ns#" about="http://peo`



“Live” box gives how the work should be attributed.

With the understanding that:

Waiver — Any of the above conditions can be **waived** if you get permission from the copyright holder.

Experiment

An experiment was conducted to measure the level of license violations upon content reuse.

- Type of license used : **Creative Commons Attribution**
- Type of content : **Flickr images**
- Sampling method : **Simple random sampling** using the **Technorati blog indexer**
- Criteria for checking Attribution : **Checking for ‘*attributionName*’ and ‘*attributionURL*’ within a reasonable scoping in the DOM.**

Creative Commons License Violations - Experimental Result 1

Creative Commons License Violations - Experimental Result 2

- Total r
- Total r
- Total r
- Total r
- Total r
- Misatt

License V



Creative Commons License Violations - Experimental Result 3

- Total number of websites tested = 70
- Total number of images in all of the websites = 466
- Total number of properly attributed images in all of the websites = 6
- **Total number of Non-Attributed Images = 439**
- Total number of images that had an error (Due to bad HTML, parsing errors, Flickr errors) = 21
- **Misattribution Percentage = 94 percent**

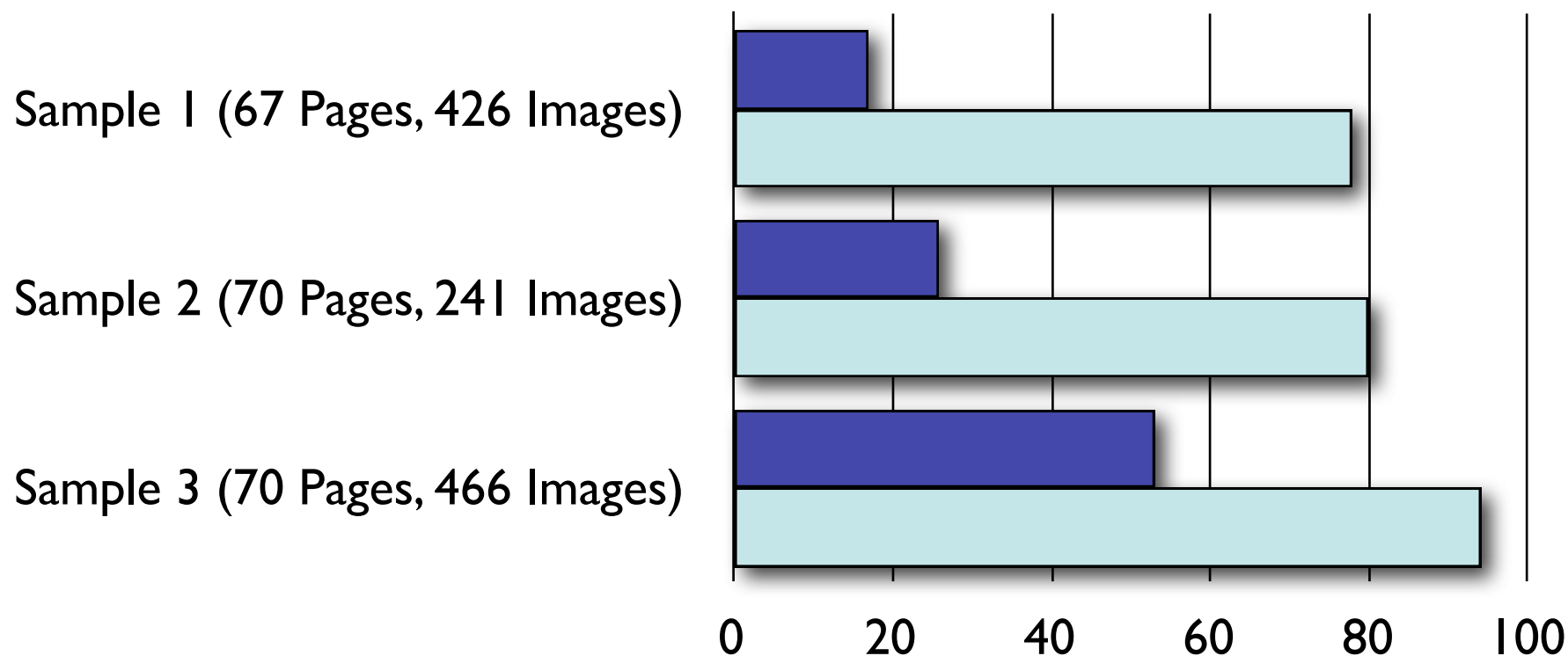
[Downloaded from ascelibrary.org by University of California, San Diego on 06/01/15. Copyright ASCE. For personal use only; all rights reserved.](#)

Non-Attributed Flickr Image	Owner	License
	kaosinthe garden	All Rights Reserved
	Jorge Cardoso	All Rights Reserved

Non-Attributed Flickr Image	Owner	License
		

10

Experiment Results



Tools to enable Policy Awareness

- Validators to verify your work
- Tools to seamlessly copy license info along with the content


Flickr CC Attribution License Violations Validator

<http://dig.csail.mit.edu/FlickrCC/validator.cgi>

Creative Commons Attribution License Violations Validator for Flickr Images

◀ ▶

+


<http://dig.csail.mit.edu/FlickrCC/validator.cgi>

↻

Q

Google





Creative Commons Attribution License Violations Validator for Flickr Images

Enter the URI of a page which has embedded Flickr images. The validator will indicate if you have not properly attributed the owner of those images (if the images are under a Creative Commons Attribution license). You can use the Attribution XHTML provided on the page to correct your errors.

URI:

Results for the page <http://dig.csail.mit.edu/FlickrCC/test/set1.html>.

Found 2 attribution license violations for the following images.

	<p>By steren.giannini Has License </p> <p>Attribution XHTML:</p> <pre><div xmlns:cc="http://creativecommons.org/ns#" about="http://farm4.static.flickr.com/3148/2732488224_aedf36e837.jpg"> steren.giannini /</pre>
	<p>By pgoeltz Has License </p>

Things that the Flickr CC Attribution License Violations Validator **cannot** handle

- Validating images that are originally from Flickr, but are downloaded and used (images that do not have a URI from Flickr)
- Correctly validating CC licensed images from Flickr of which the rights do not belong to the uploader

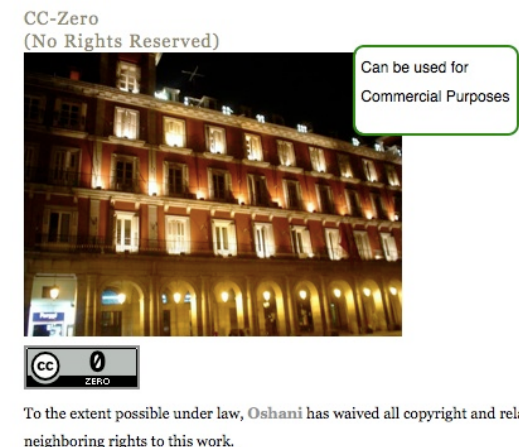
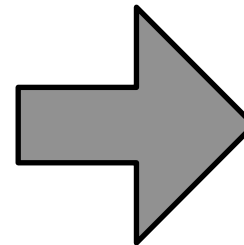
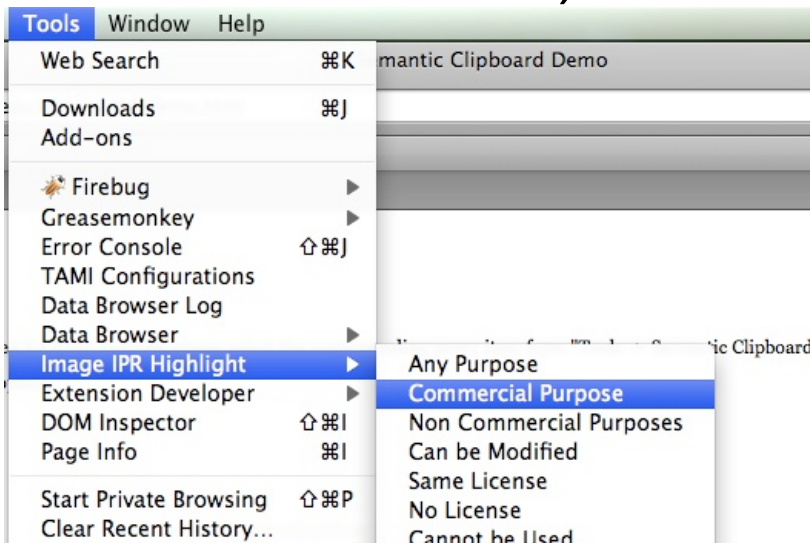
Semantic Clipboard

(Integrated with Tabulator Firefox Extension)

- Shows if an image can be copied or not (based on the license it is under)



- Can use this tool to see which images can be used for a particular purpose (for e.g. pick out the images that can be used for a commercial use)

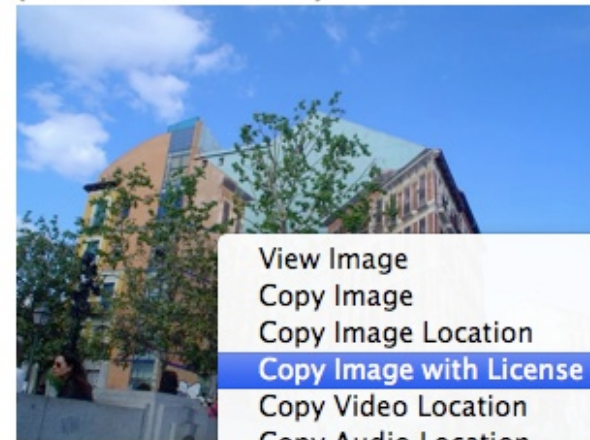


Semantic Clipboard

(Integrated with Tabulator)

- **Copy an image with the license**
 - Scrape the License RDFa
 - Construct the Attribution XHTML
 - Paste into any application

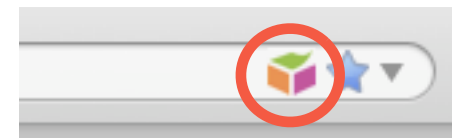
CC-BY
(Use with attribution)



This work by Osha
Unported License

- View Image
- Copy Image
- Copy Image Location
- Copy Image with License**
- Copy Video Location
- Copy Audio Location
- Save Image As...
- Send Image...
- Set As Desktop Background...
- Block Images from people.csail.mi...
- Properties
- Inspect Element

- **Side Effect:**
 - Extracts RDFa in a page that a user browses



Things that the Semantic Clipboard ***cannot*** handle

- Images in which the license metadata is not expressed in RDFa
- Copying of media types other than images (but it can be easily extended to other types of media as long as the licenses are expressed in RDFa)
- License conflicts in the destination document

Questions ?

oshani@csail.mit.edu