



Policy Aware Content Reuse on the Web

Oshani Seneviratne, Lalana Kagal and Tim Berners-Lee Decentralized Information Group, MIT

International Semantic Web Conference October 27, 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 (Both figures are as of June 2009)



- 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 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 ¹	
	Allow modifications of your work?	
	• Yes •	
-	O Yes, as long as others share alike 0	
	○ No 	
Jurisdiction of your license		
	United States 💠	





Anatomy of a CC License

Human readable form





This work by <u>John Doe</u> is licensed under a <u>Creative Commons Attribution 3.0 United States License</u>. Based on a work at http://www.example-source.com.

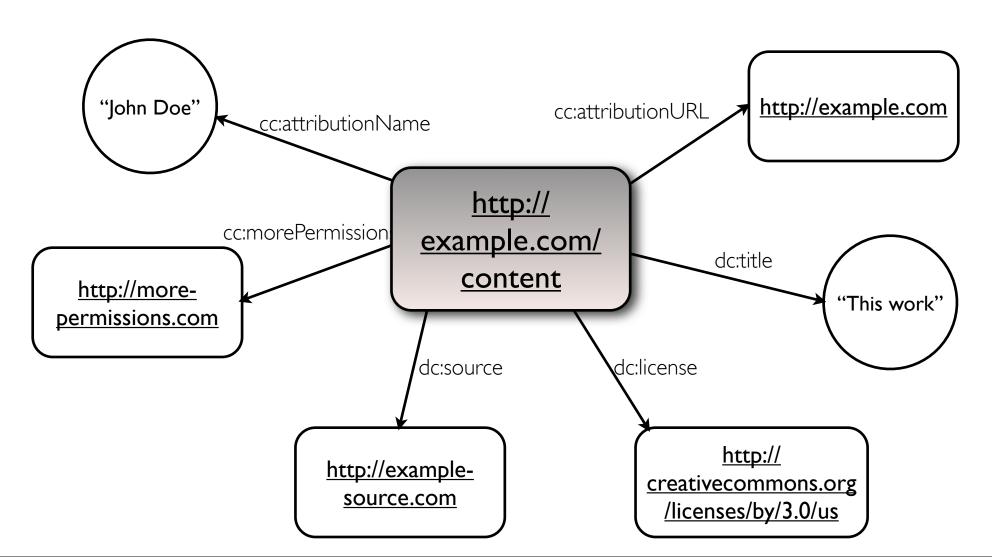
Permissions beyond the scope of this license may be available at http://morepermissions.com





Anatomy of a CC License

Things conveyed in the underlying RDF







Value

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=
▼ <a rel="license" href="http://creativecommons.org/licenses/by/3.0/us/">
    <img alt="Creative Commons License" style="border-width:0" src="http://i.creativecommons.org/l/by/3.0/us/88x31.png ">
  </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>
  <hr>
 " Permissions beyond the scope of this license may be available at "
  <a href="http://more-permssions.com" rel="cc:morePermissions">http://morepermssions.com</a>
                                                                                                               Legend
</div>
                                                                                                                  Attribute
```





CC Deeds using RDFa and Javascript



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



© creative commons

Attribution 3.0 United States



Under the following conditions:



"Live" box gives how the work should be attributed.





Are users aware of these tool and techniques?

Apparently Not!
We found 78% - 94% license violations on the Web.





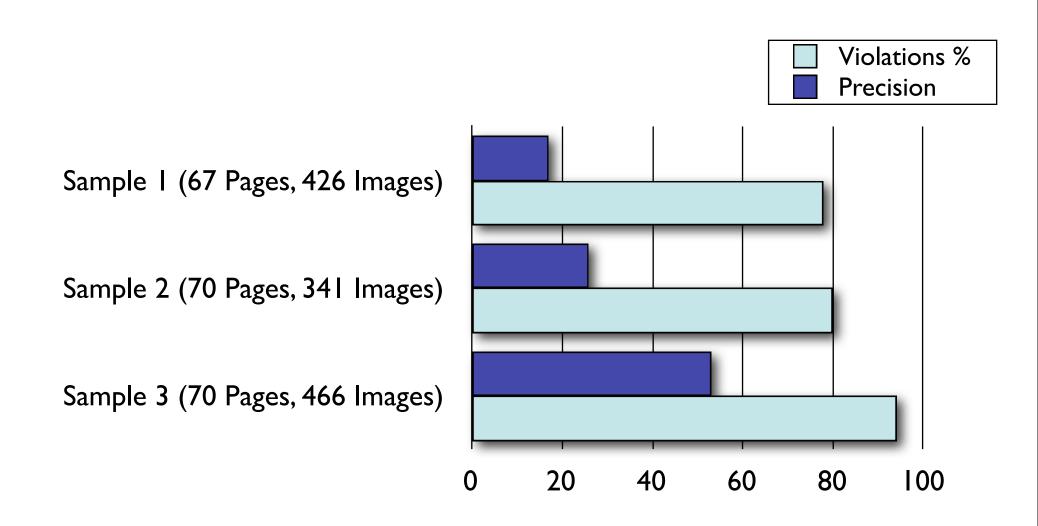
Experiment

- 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.





Experiment Results







Experiment Results

http://dig.csail.mit.edu/2008/WSRI-Exchange/results

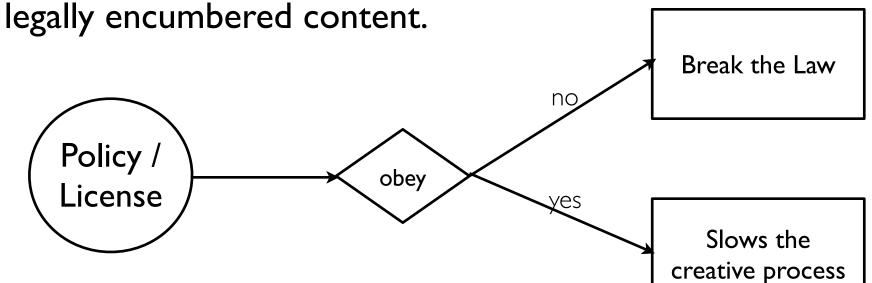






Problems with Content Reuse

A potential legal problem arises when combining one or more



- License information is not always visible or available.
- Most users do not know about licenses or they may be too lazy to be license compliant.





Tools to enable Policy Awareness

- Validators to verify your work
- Tools to seamlessly copy license info





Flickr CC Attribution License Violations Validator http://dig.csail.mit.edu/FlickrCC/validator.cgi

	Commons Attribution License Violations Validator for Flickr Images	
→ + 6 http://d	dig.csail.mit.edu/FlickrCC/validator.cgi	
Images Enter the URI of a page	mons Attribution License Violations Validator for Flickr which has embedded Flickr images. The validator will indicate if you have not properly attributed the (if the images are under a Creative Commons Attribution license). You can use the Attribution XHTML correct your errors.	
Cear Results Results for the page http://dig.csail.mit.edu/FlickrCC/test/set1.html . Found 2 attribution license violations for the following images.		
	By <u>steren.giannini</u> Has License to 1	
	Attribution XHTML: <div about="http://farm4.static.flickr.com/3148/2732488224_aedf36e837.jpg" xmlns:cc="http://creativecommons.org/ns#"> steren.giannini / //</div>	
	By <u>pgoeltz</u> Has License ⓒ ���⊜	





Demo

http://dig.csail.mit.edu/2009/Talks/1027-iswc-os/flickrcc_demo.mov





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

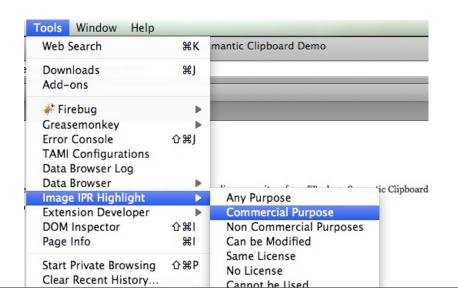


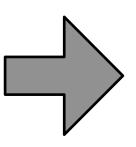


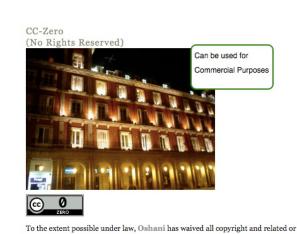
Semantic Clipboard

• 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)







neighboring rights to this work.





Semantic Clipboard

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







Demo

http://dig.csail.mit.edu/2009/Talks/1027-iswc-os/semclip_demo_new.mov



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)





Related Work

- Attributor, PicScout for finding out copyright violations on the Web
- CC License Syntax Validation service
 - http://validator.creativecommons.org
- MozCC, a FF extension that displays CC rights information,
 Nathan Yergler
- XHTML documents with inline policy provenance, Harvey Jones





Conclusion

- Experiment results prove that there are many license violations on the Web.
 - http://dig.csail.mit.edu/2008/WSRI-Exchange/results
- FlickrCC Validator can be used to validate a users' document to keep an honest person honest.
 - http://dig.csail.mit.edu/FlickrCC/validator.cgi
- Semantic Clipboard can be used to copy attribution details along with the content.
 - http://dig.csail.mit.edu/2009/Clipboard





Questions?

oshani@csail.mit.edu

Redistribution License:

