

# 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 **You Tube**



(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 **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 

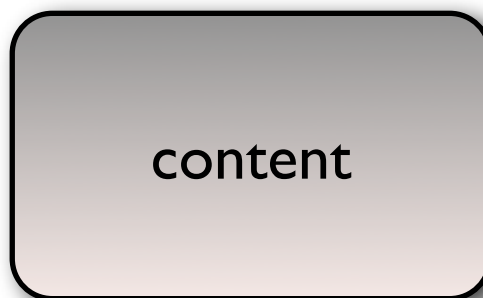
No 

Jurisdiction of your license 

United States 

# Anatomy of a CC License

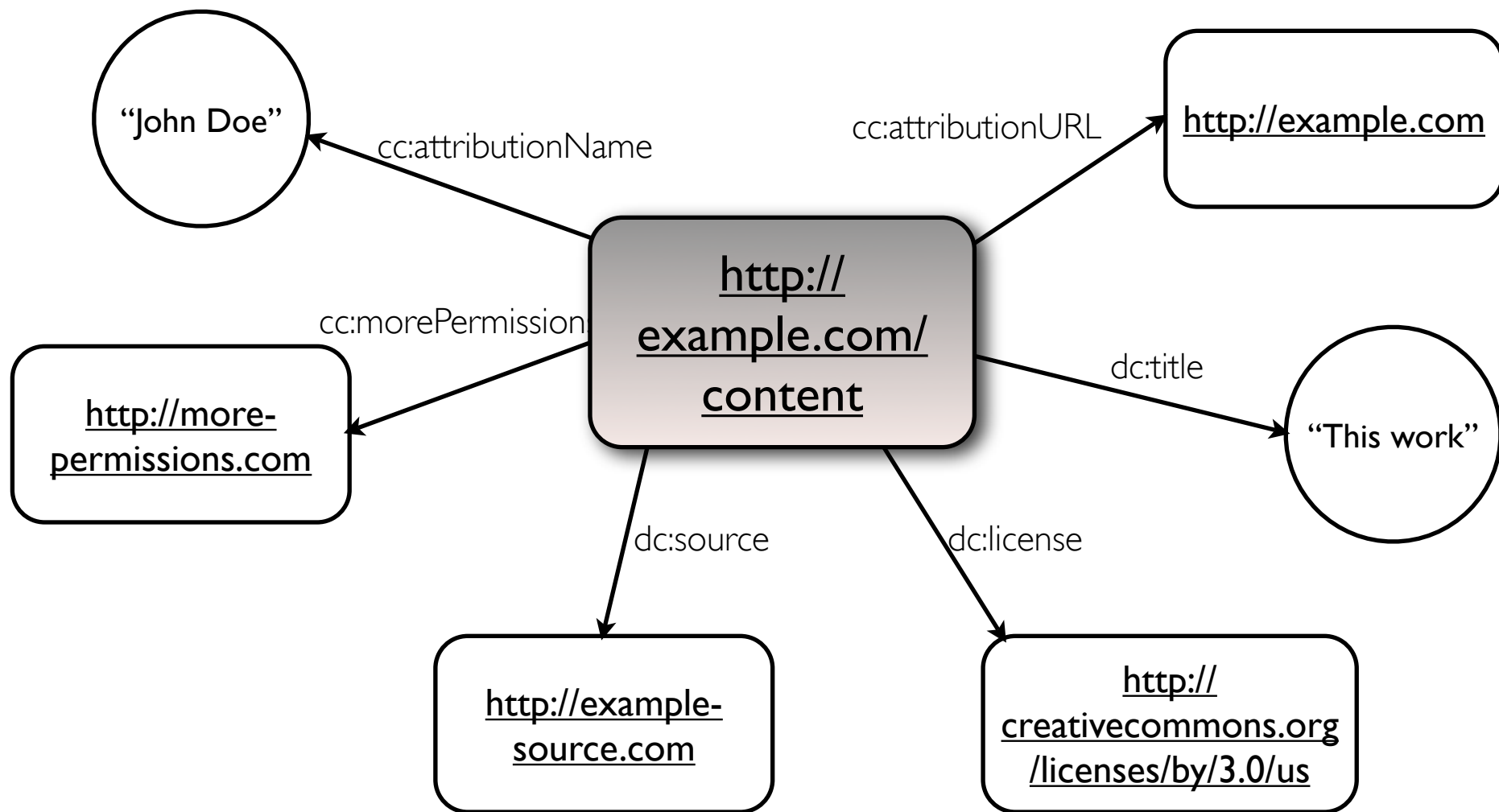
Human readable form



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

Things conveyed in the underlying RDF



# 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/" >
    
  </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

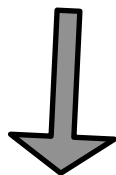
Attribute

Value


# 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>




## 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](#)



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.



Are users aware of these tool and techniques?

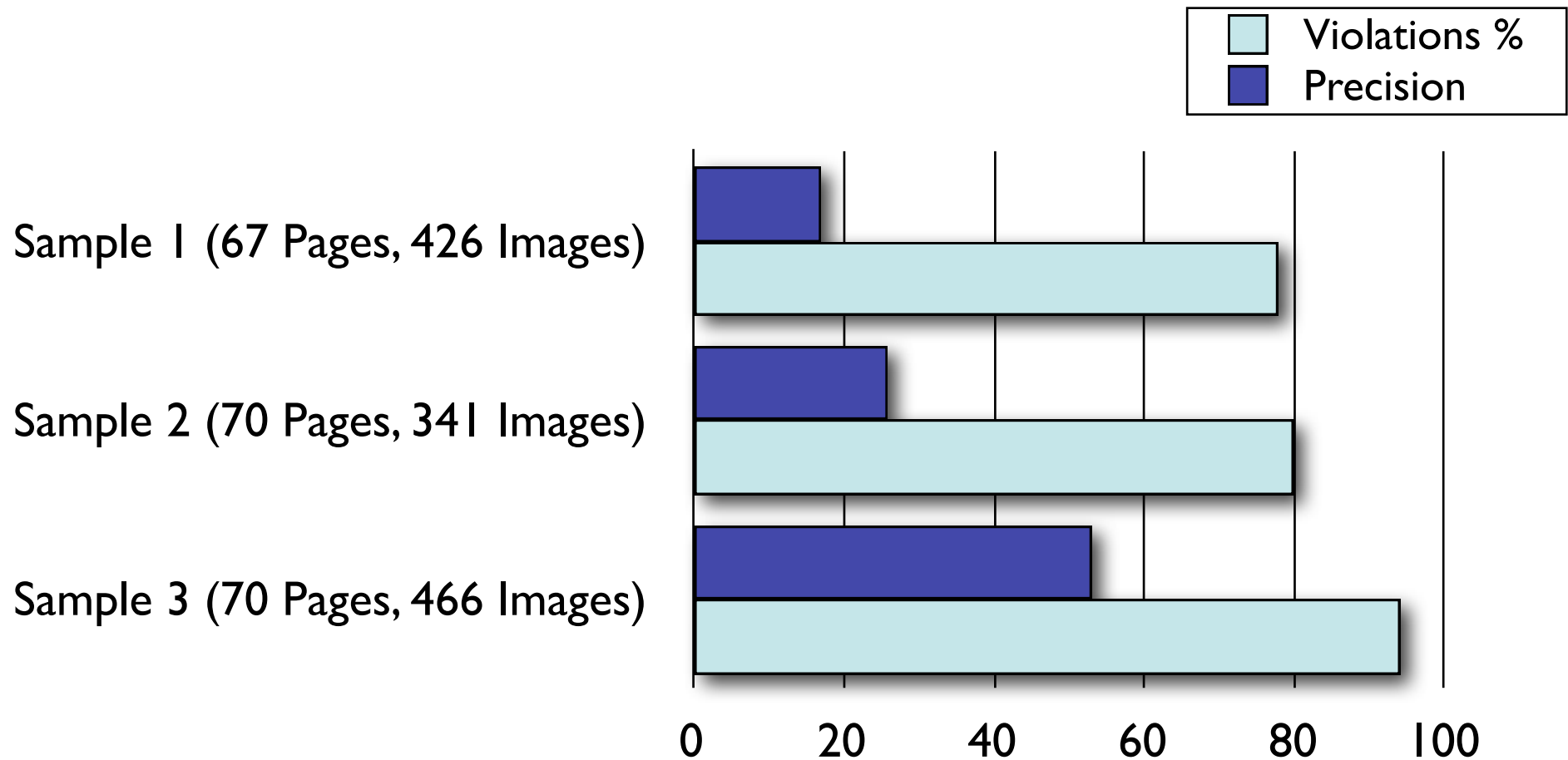
Apparently Not!

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/WWSRI-Exchange/results>

## Creative Commons License Violations - Experimental Result 1

### Statistics

- Total
- Total
- Total
- Total
- Total
- **Mis**

### License V



## Creative Commons License Violations - Experimental Result 2

### Statistics

- Total r
- Total r
- Total r
- Total r
- Total r
- **MisatB**

### License V



## Creative Commons License Violations - Experimental Result 3

### Statistics

- 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**

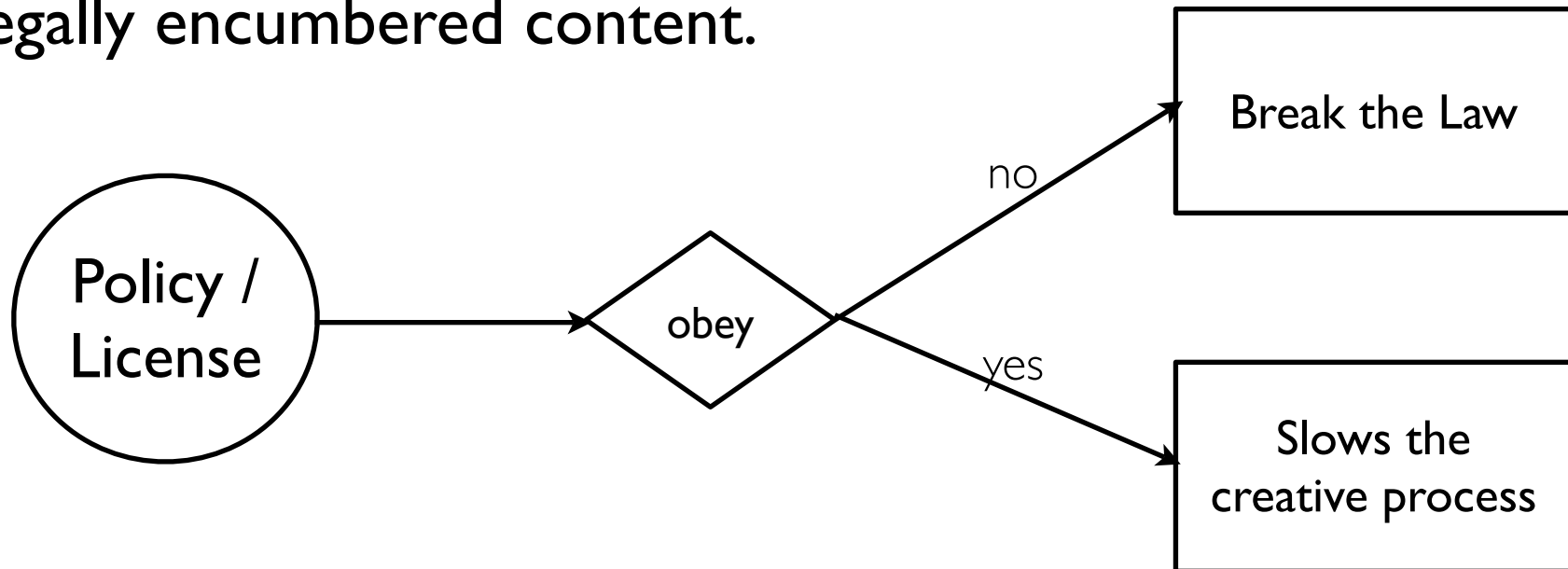
## License Violations Detected in Each Individual Sites

Non-Attributed Flickr Image	Owner	License
	<a href="#">kaosinthegarden</a>	All Rights Reserved
	<a href="#">Jorge_Cardoso</a>	All Rights Reserved

Non-Attributed Flickr Image	Owner	License
		

# Problems with Content Reuse

- A potential legal problem arises when combining one or more legally encumbered content.



- 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>

Creative Commons Attribution License Violations Validator for Flickr Images

← →
+
http://dig.csail.mit.edu/FlickrCC/validator.cgi
↻

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:  Validate!


---

Clear Results

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

Found 2 attribution license violations for the following images.

---




By [steren.giannini](#)  
 Has License

*Attribution XHTML:*

```

            <div xmlns:cc="http://creativecommons.org/ns#"
            about="http://farm4.static.flickr.com/3148/2732488224_aedf36e837.jpg">
            <a rel="cc:attributionURL"
            href="http://www.flickr.com/people/steren.giannini">steren.giannini</a> /
            
```

---



By [pgoeltz](#)  
 Has License



# Demo

A screenshot of a web browser window. The title bar reads "Flickr CC Validator Demo". The address bar shows the URL "http://people.csail.mit.edu/oshani/iswc\_demo/flickrcc.html". The browser has three tabs open: "Flickr CC Validator Demo", "Flickr: Creative Commons", and "Creative Commons Attribution...". The main content area of the browser is a light green page with the heading "Flickr CC Validator Demo" at the top. The rest of the page is blank. A mouse cursor is visible in the center of the page.

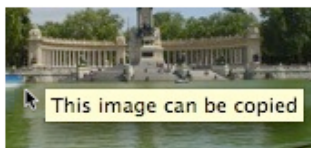


# Things that the Flickr CC Attribution License Violations Validator **cannot** handle

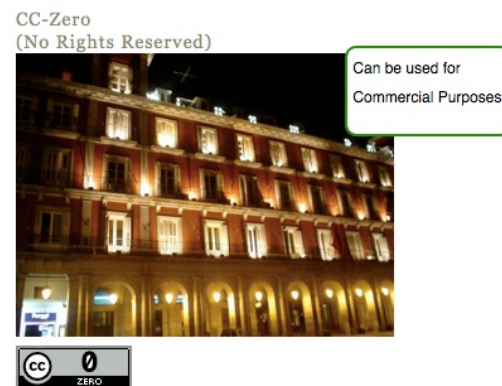
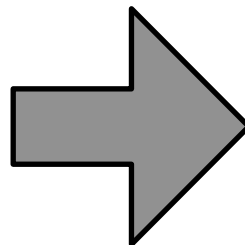
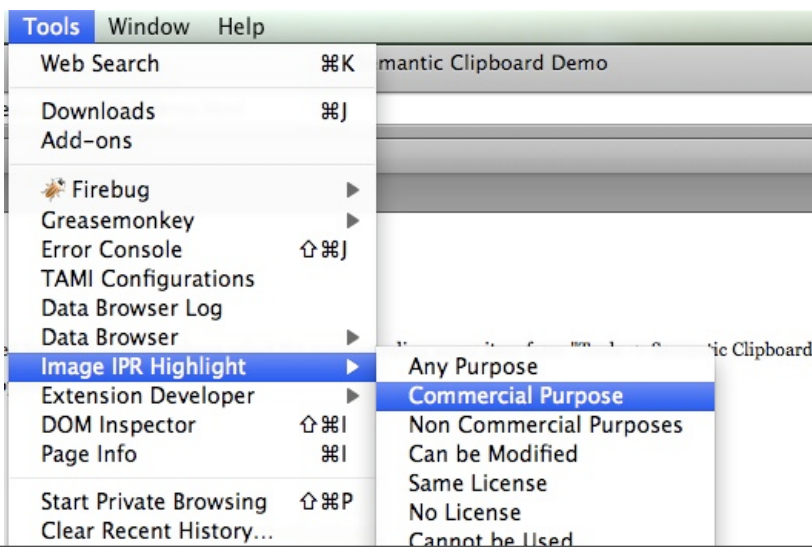
- Validating images that are originally from Flickr, but are downloaded and used (images that do not have a URI from Flickr)
- Correctly validating CC licensed images from Flickr of which the rights do not belong to the uploader

# 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)



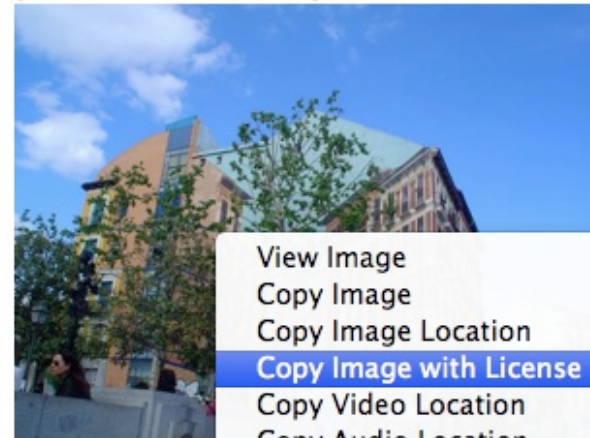
To the extent possible under law, Oshani has waived all copyright and related or neighboring rights to this work.

# Semantic Clipboard

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

CC-BY

(Use with attribution)



This work by Oshar

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

on 3.0


# Demo

Semantic Clipboard Demo


http://people.csail.mit.edu/oshani/SM/demo.html

Semantic Clipboard Demo

No License  
(All Rights Reserved)




CC-BY-NC  
(Cannot be used for commercial uses)



3.0 Unported License.


CC-BY-ND  
(Cannot be modified)



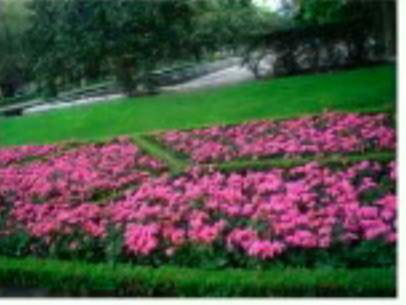
This work by Oshani is licensed under a Creative Commons Attribution Non-Commercial 3.0 Unported License.

This work by Oshani is licensed under a Creative Commons Attribution No-Derivative 3.0 Unported License.


CC-BY-NC-ND  
(No commercial use & Cannot be modified)



CC-BY-NC-SA  
(No commercial use & Same License)



CC-BY-NC-SA  
(Older CC license)



This work by Oshani is licensed under a Creative Commons Attribution Non-Commercial No-Derivative 3.0 Unported License.

This work by Oshani is licensed under a Creative Commons Attribution Non-Commercial Share-Alike 3.0 Unported License.

This work by Oshani is licensed under a Creative Commons Attribution Non-Commercial Share-Alike 2.5 Generic License.

# 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/WWSRI-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](mailto:oshani@csail.mit.edu)

Redistribution License:

