

# Policy Aware Content Reuse

---

Oshani Seneviratne



Social Web Incubator Group  
April 14<sup>th</sup>, 2010





# Outline

---

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy



# Outline

---

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy

# Content on the Web

---

- There is lot of content on the Web
  - 3.6 billion images on **flickr**
  - 20 hours of video uploaded every minute on **YouTube**  
Broadcast Yourself™
  - Over 1 trillion web pages on the web  
(source: <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>)
- Content reuse is important
  - Prevents redundant work
  - Promotes creativity



# creative commons Licenses

- De-facto license for open content distribution
- 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
- Large community

# creative commons Licenses

- There are four different categories (clauses)

- Attribution (BY)
- Share-Alike (SA)
- Non-Commercial use (NC)
- No-Derivatives (ND)

- These can be mixed in any combination to generate custom licenses.

- Attribution
- Attribution-NoDerivs
- Attribution-NonCommercial-NoDerivs
- Attribution-NonCommercial
- Attribution-NonCommercial-ShareAlike
- Attribution-ShareAlike



# Most Search Engines support CC

**Google** flowers  [Advanced Search](#)

Search only pages that are free to use, share or modify, even commercially  Search the Web

---

**YAHOO!** Web Images Video Local Shopping More ▾

flowers  Options ▾

Creative Commons:  for commercial use  to modify

**SEARCH NOW!**

**CC GET MEDIA** flowers

Options



# Most Content Distribution Sites support CC

**flickr**® from YAHOO! You aren't signed in

[Home](#) [The Tour](#) [Sign Up](#) [Explore](#) ▾

---

**Search** Photos | Groups | People

Everyone's Uploads ▾  [SEARCH](#) [Full Text](#) | [Tags Only](#) [Advanced Search](#)

⊕ Showing Creative Commons-licensed content for commercial use, adaptation, modification or building upon ×

**jamendo** open your ears [Music](#) [Selections](#) [Community](#) [Jamendo PRO](#) [About](#)

Home > Music > Tags > flowers

You like "flowers" ? Here is some of the best "flowers" on jamendo. You can listen, download and share it, it's free.

All the albums | Short albums | Full-length albums

Browsing 31082 albums 1 2 3 4 5 > [1555]

Order by  ▾

Search by tag

By country  ▾

Find content I can use for commercial purposes.  Find content I can modify, adapt or build upon.

[Cancel the filter](#)

**blip.tv** [Browse](#) ▾ [Dashboard](#) ▾ [Upload](#) ▾ [Help](#) ▾  [Search blip.tv](#)

**Choosing a Camera**

**Lighting**

**EPISODES** | **SHOWS**

Displaying popular episodes licensed under licenses that permit commerce and modification [Start Over](#)

<a href="#">Browse by ...</a>	<a href="#">My favorite shows</a>	<a href="#">Options</a>	<a href="#">Show explicit content</a>
-------------------------------	-----------------------------------	-------------------------	---------------------------------------







# Generating a Creative Commons License


- 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 

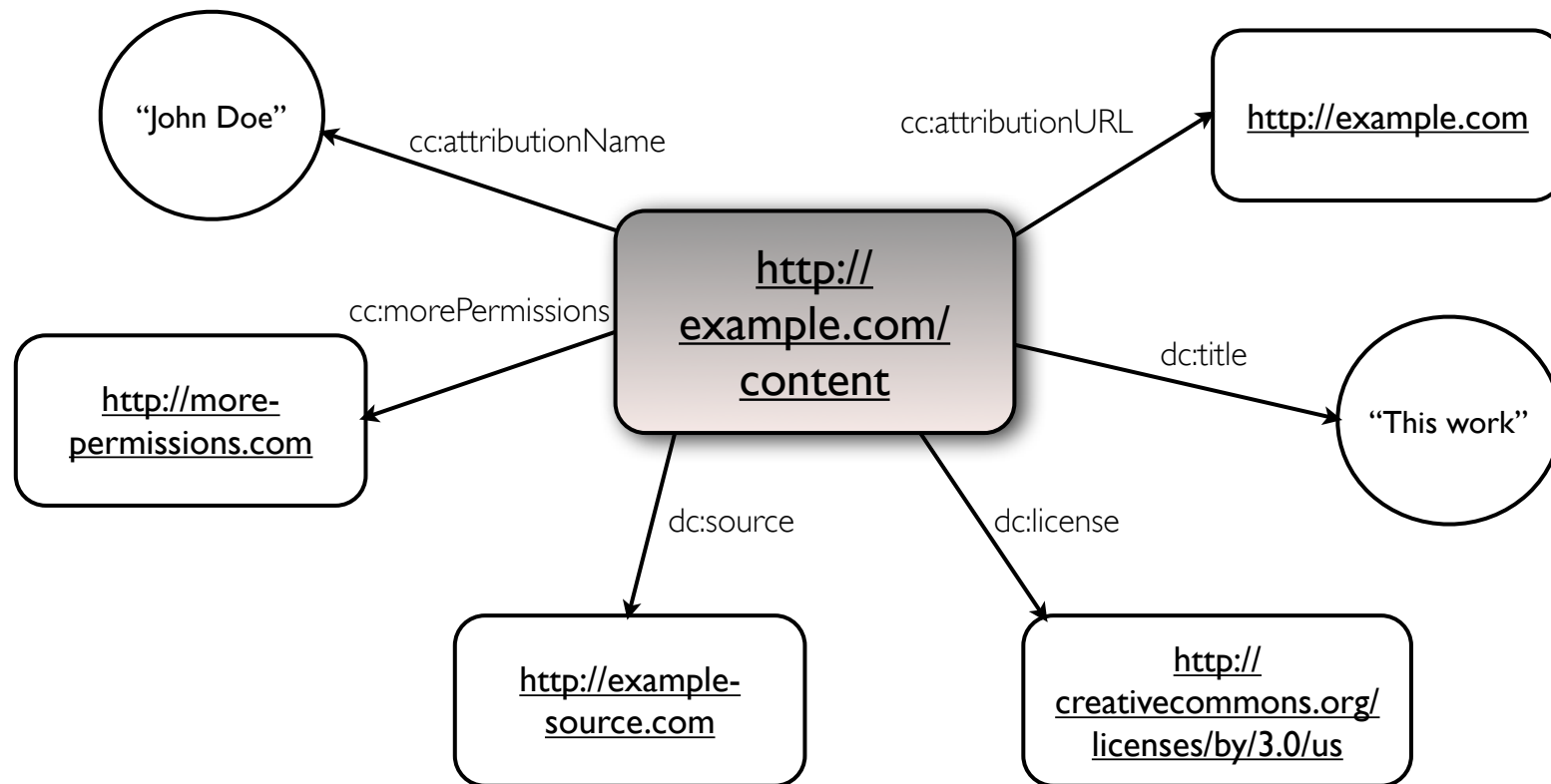
No 

Jurisdiction of your license 

United States 

# Anatomy of a CC License

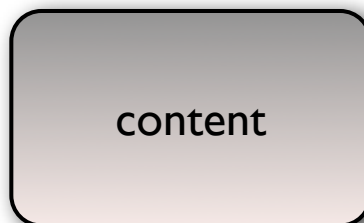
## Things that are conveyed in a license



# Anatomy of a CC License

---

## Human Readable Form

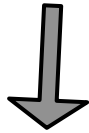


This work by [John Doe](#) is licensed under a [Creative Commons Attribution 3.0 United States License](#).  
Based on a work at <http://www.example-source.com>.  
Permissions beyond the scope of this license may be available at <http://morepermissions.com>

# CC Deed



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.

## With the understanding that:

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



---

Are users aware of these licenses, and tools available for content reuse?



---

Are users aware of these licenses, and tools available for content reuse?

Apparently not!

We found **78% - 94%** attribution license violations with Flickr images on the Web.

# Experiment

---

Type of content used: **Flickr images**

Why Images?

- Most common type of web content reused in blogs

Why Flickr?

- Flickr has over 100 million CC-licensed images
- Flickr API supports querying of CC license information
- Flickr images have a fixed image URI:

`http://farm{farm-id}.static.flickr.com/{server-id}/{id}_{secret}.jpg`

# Experiment

---

## Samples - **Technorati blog indexer to generate the samples**

### Why Technorati?

- Technorati blog indexer crawls and indexes over 60 million weblogs
- Possible to obtain weblogs that link to a particular URL
- The 'authority rank' for weblogs changes with time

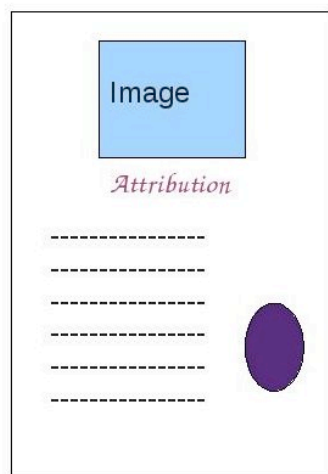


# Experiment

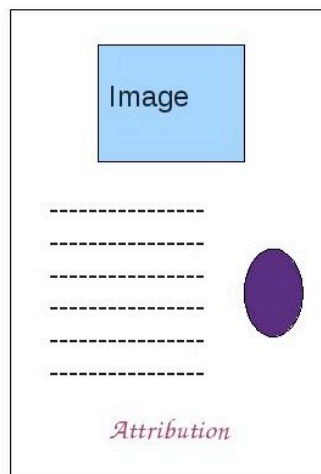
## Type of license used - **Criteria for checking Attribution**

Checking one of these within reasonable scoping in the DOM where the image is embedded:

- Flickr username or full name (*cc:attributionName*)
- Flickr user profile URI (*cc:attributionURL*)



Attribution Example 1



Attribution Example 2



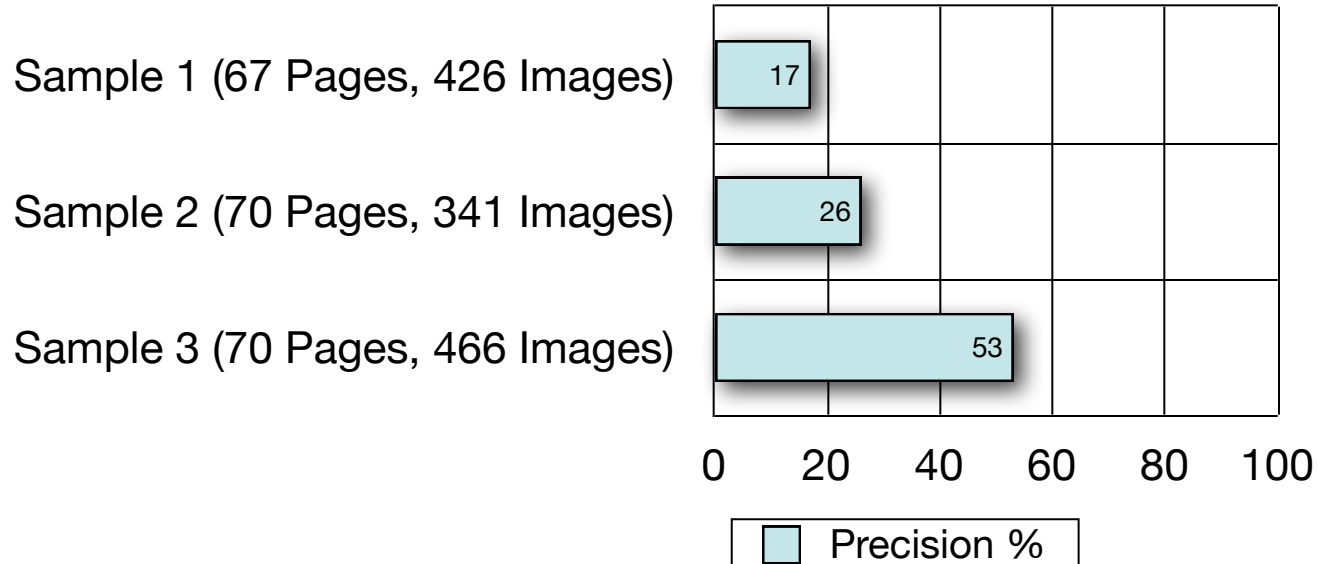
# Results

$$\text{Raw Violations} = \frac{\text{Total \# of non/mis-attributed images}}{\text{Total \# of images}}$$



# Results

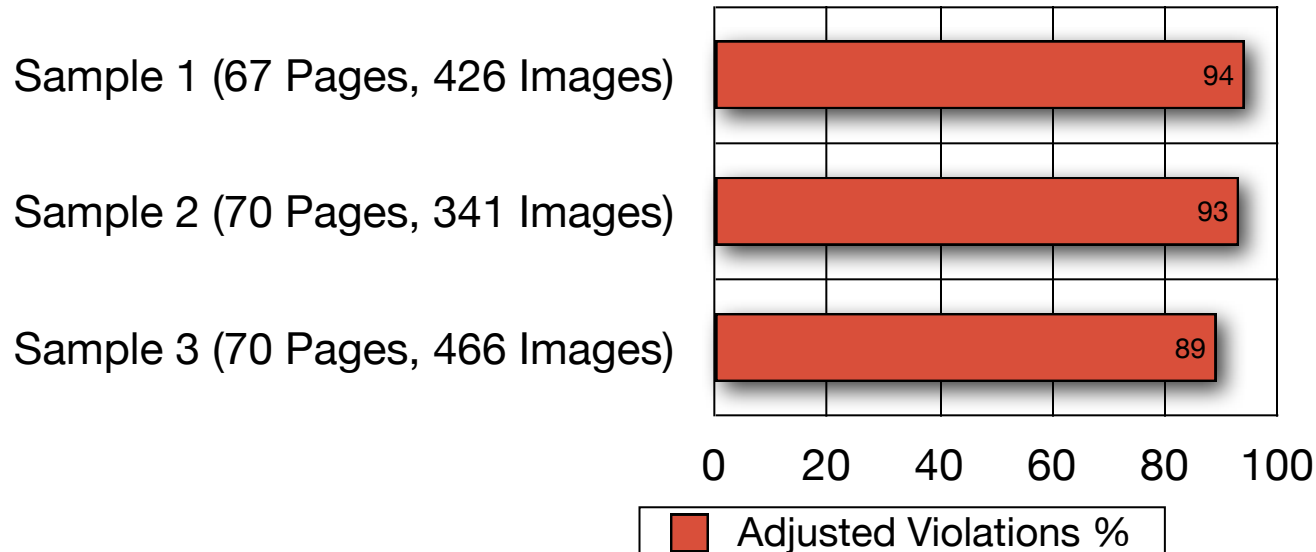
$$\text{Precision} = \frac{\text{Total \# of violations} - \text{Total \# of images owned by the user}}{\text{Total \# of violations}}$$



# Results

Adjusted Violations =

$$\frac{\text{Total \# of images that are not owned by the reuser and that are not attributed}}{\text{Total \# of images that not are owned by the reuser}}$$



# Results

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

**Creative Commons License Violations - Experimental Result 1**

**Statistics**

- Total
- Total
- Total
- Total
- Misat

**Creative Commons License Violations - Experimental Result 2**

**Statistics**

- Total
- Total
- Total
- Total
- Total
- Misat

**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**

[http://www.kaosinthe.com/2008/08/08/...](#)

Non-Attributed Flickr Image	Owner	License
	<a href="#">kaosinthe garden</a>	All Rights Reserved
	<a href="#">Jorge Cardoso</a>	All Rights Reserved

[http://www.kaosinthe.com/2008/08/08/...](#)

Non-Attributed Flickr Image	Owner	License



# Outline

---

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy


# Flickr CC Attribution License Violations Validator


Creative Commons Attribution License Violations Validator for Flickr Images

URI:

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 

*Attribution XHTML:*

```
<div xmlns:cc="http://creativecommons.org/ns#"
about="http://farm4.static.flickr.com/3428/3894160803_842929aaf1.jpg">
<a rel="cc:attributionURL"
href="http://www.flickr.com/people/pgoeltz">pgoeltz</a> / <a
```

The owner/source of the following image was attributed. But use the proper XHTML provided.



By [CornGuo](#)  
Has License 

*Attribution XHTML:*

```
<div xmlns:cc="http://creativecommons.org/ns#"
about="http://farm3.static.flickr.com/2662/3894633614_2cf2c25015.jpg">
<a rel="cc:attributionURL"
```



# Limitations

---

- 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





# Flickr CC Attribution License Violations Validator

---

Demo:

[http://people.csail.mit.edu/oshani/talks/flickrcc\\_demo.mov](http://people.csail.mit.edu/oshani/talks/flickrcc_demo.mov)

Try it out:

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



# Outline

---

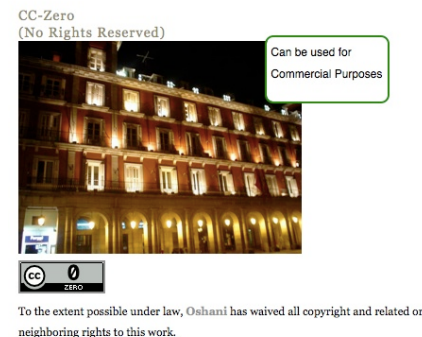
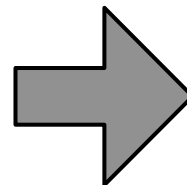
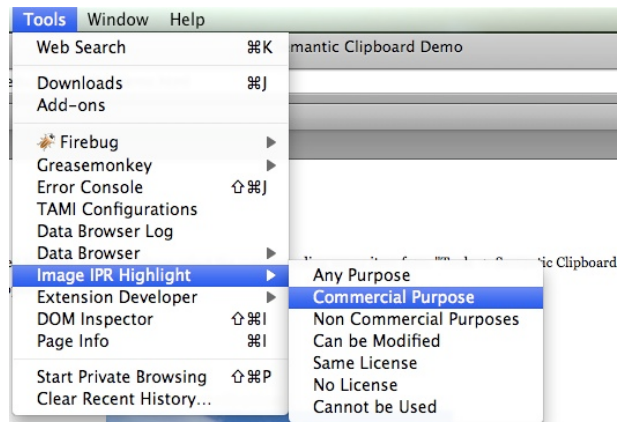
- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy

# Semantic Clipboard

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

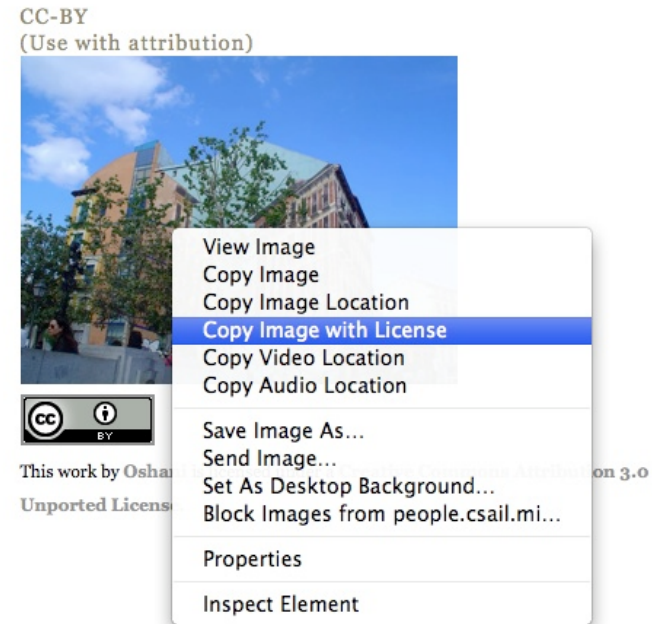


- Can use this tool to see which images can be used for a particular purpose (for e.g. pick out the images that can be used for a commercial use)



# Semantic Clipboard

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



# Limitations

---

- 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)



# Semantic Clipboard

---

Demo:

[http://people.csail.mit.edu/oshani/talks/semclip\\_demo.mov](http://people.csail.mit.edu/oshani/talks/semclip_demo.mov)

Try it out:

Install Tabulator Firefox Extension from <http://dig.csail.mit.edu/2007/tab/release/unofficial/tabulator-latest.xpi>



# Outline

---

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy (Work by Ted Kang and Lalana Kagal)



# Social Networks & Privacy

- Traditional Method of Access Control:
  - Build a big fence around private information and hope nothing gets out.

The screenshot shows the Facebook Privacy Settings page for a user's profile information. The page is titled "Privacy Settings > Profile Information" and includes a "Back to Privacy" button and a "Preview My Profile..." button. The settings are organized into several sections, each with a privacy dropdown menu set to "Only Friends":

- About me**: About Me refers to the About Me description in your profile. (Only Friends)
- Personal Info**: Interests, Activities, Favorites. (Only Friends)
- Birthday**: Birth date and Year. (Only Friends)
- Religious and Political Views**: (Only Friends)
- Family and Relationship**: Family Members, Relationship Status, Interested In, and Looking For. (Only Friends)
- Education and Work**: Schools, Colleges and Workplaces. (Only Friends)
- Photos and Videos of Me**: Photos and Videos you've been tagged in. (Only Friends)
- Photo Albums**: (Edit Settings)
- Posts by Me**: Default setting for Status Updates, Links, Notes, Photos, and Videos you post. (Only Friends)
- Allow friends to post on my Wall**: (Checked) Friends can post on my Wall
- Posts by Friends**: Control who can see posts by your friends on your profile. (Only Friends)
- Comments on Posts**: Control who can comment on posts you create. (Only Friends)

At the bottom, there is a link to "Application Settings Page" and a footer with "Facebook © 2010 English (US)" and "About Advertising Developers Careers Terms • Find Friends Privacy Mobile Help Center".



# Social Networks & Privacy

---

- What happens when the data gets out by being:
  - Copied and reposted
  - Inference (for example project “gaydar”)
  - Policy changes
  
- The problem: What happens when the data gets out?



# Social Networks & Privacy

---

- ◆ Is there an alternative to enforcing privacy policies through restricted access?
- ◆ Conform to existing policies by making users **aware of the usage restrictions** associated with the data



# Social Networks & Privacy

---

- ◆ “Respect My Privacy” is supported across three platforms:
  - Facebook
  - OpenSocial
  - Tabulator Semantic Web browser
  
- ◆ These applications enable users to **specify** privacy preferences for their data and also **display privacy-annotated data** prominently enabling other users to easily recognize and conform to these preferences

# RMP in Facebook/OpenSocial

◆ Five RMP restrictions:

- no-commercial
- no-depiction
- no-employment
- no-financial
- no-medical



# RMP in Facebook

- RMP application for Facebook:
  - <http://apps.facebook.com/respectmyprivacy/>

The screenshot shows a Facebook profile for Ted Kang. The profile includes a cover photo, a profile picture, and a navigation bar with tabs for Wall, Info, Photos, and Boxes. Below the navigation bar is a text input field for "What's on your mind?" with a "Share" button. The main content area displays a list of posts from friends, including Luke Peasley, Jennifer Chou, Jefferson Shin, and Steven Wysocky. A "Middle of Profile Omitted" section is visible, followed by a post from Albert Chang and a post from Jonathan Fu. On the left side of the profile, there is a "Respect My Privacy" application interface. This interface includes a "View Photos of Me (50)" link, an "Edit My Profile" link, a text input field with "hi i'm ted.", an "Information" section with details like "Networks: MIT '08", "Relationship Status: Single", and "Birthday: June 11, 1986". The RMP application also features a "Respect My Privacy" button with a "Learn more ..." link and a warning message: "The information in this profile may not be used for employment purposes."

# RMP on the Semantic Web

---







- ◆ Transparency handled. But how about automated accountability mechanisms?
- ◆ Our solution: Move to decentralized social networks
- ◆ The FOAF Converter developed by Matthew Rowe:
  - Translates Facebook profile pages to Semantic Web data using the Friend of A Friend (FOAF) ontology
- ◆ The five RMP restrictions are represented as ontological data
- ◆ The **restricts** property allows restrictions to be attached to any resource
- ◆ Additional RMP restrictions are added to the converted FOAF data

# Tabulator

---

- ◆ **Goal:** Integrate the Semantic Web directly into the browser
- ◆ Use panes for viewing specific types of data
  - Triple, rule, justification, social panes, microblogging etc.
- ◆ Navigate by clicking through links to various resources
- ◆ Edit resources in-place via SPARUL (SPARQL Update)
- ◆ Available as a Firefox extension




# Tabulator: Triple/Data Pane


▼ Tim Berners-Lee's FOAF file      

W3C Standards and Technical Reports	Title	W3C Standards and Technical Reports	
Tim Berners-Lee's FOAF file	license	<a href="http://creativecommons.org/licenses/by-nc/3.0/">http://creativecommons.org/licenses/by-nc/3.0/</a>	
	Title	Tim Berners-Lee's FOAF file	
	type	<a href="http://xmllns.com/foaf/0.1/PersonalProfileDocument">http://xmllns.com/foaf/0.1/PersonalProfileDocument</a>	
	maker	Timothy Berners-Lee	
	primary Topic	Timothy Berners-Lee	
Coralie Mercier	type	<a href="http://xmllns.com/foaf/0.1/Person">http://xmllns.com/foaf/0.1/Person</a>	
	seeAlso	<a href="http://www.koalie.net/foaf.rdf">http://www.koalie.net/foaf.rdf</a>	
	mbox	<a href="mailto:coralie@w3.org">mailto:coralie@w3.org</a>	
	name	Coralie Mercier	
Dean Jackson	type	<a href="http://xmllns.com/foaf/0.1/Person">http://xmllns.com/foaf/0.1/Person</a>	
	seeAlso	<a href="http://www.grorg.org/dean/foaf.rdf">http://www.grorg.org/dean/foaf.rdf</a>	
	homepage	<a href="http://www.grorg.org/dean/">http://www.grorg.org/dean/</a>	
	mbox	<a href="mailto:dean@w3.org">mailto:dean@w3.org</a>	
		<a href="mailto:dino@grorg.org">mailto:dino@grorg.org</a>	
	mbox sha1sum	6de4ff27ef927b9ba21ccc88257e41a2d7e7d293	
name	Dean Jackson		
Edd Dumbill	type	<a href="http://xmllns.com/foaf/0.1/Person">http://xmllns.com/foaf/0.1/Person</a>	
	seeAlso	<a href="http://heddley.com/edd/foaf.rdf">http://heddley.com/edd/foaf.rdf</a>	
	homepage	<a href="http://heddley.com/edd/">http://heddley.com/edd/</a>	
	mbox	<a href="mailto:edd@usefulinc.com">mailto:edd@usefulinc.com</a>	
		<a href="mailto:edd+xml.com">mailto:edd+xml.com</a>	
		<a href="mailto:edd+xmlhack.com">mailto:edd+xmlhack.com</a>	
name	Edd Dumbill		
nick	edd		
Henry Story	name	Henry Story	
<a href="http://dbpedia.org/resource/Eli_Lilly_and_Company">http://dbpedia.org/resource/Eli_Lilly_and_Company</a>	member	mbox	<a href="mailto:susie.stephens@gmail.com">mailto:susie.stephens@gmail.com</a>
		name	Susie Stephens
James Martin	type	<a href="http://xmllns.com/foaf/0.1/Person">http://xmllns.com/foaf/0.1/Person</a>	



# Tabulator: Social Pane


▼ Dan Brickley   



This is you











**Basic Information**

<a href="#">danbri homepage</a>
<a href="#">my.opera homepage</a>
<a href="#">danbri.livejournal weblog</a>
<a href="#">my.opera weblog</a>

Openid  <http://danbri.livejournal.com/>


Nickname danbri  
danbri

**Acquaintances**

Aaron Swartz
Amy van der Hiel
Art Barstow
<a href="#">Benjamin Joffe</a> 
<a href="#">Charles McCathieNevile</a> 
Damian Steer
Dan Connolly 
Dave Beckett
Dean Jackson 
Edd Dumbill 
<a href="#">Eddie Lopez</a> 
Eric Miller 
<a href="#">Eva Méndez</a> 
<a href="#">Gregory J. Rosmaita</a> 
 Jan Grant
Jim Ley
Joe Brickley

Tip: Do you have [web ID](#)?

[Make or set A Web ID](#)

SeeAlso  <http://foaf.qdos.com/reverse/?path=http://danbri.org/foaf.rdf#danbri>

# RMP in the Tabulator Social Pane

- ◆ Allows users to attach RMP restrictions to their FOAF files, and displays the accompanying RMP icon.

The screenshot shows a social pane for a user named Ted Kang. On the left, there is a profile picture of a woman. Below it, the text reads: "You and Ted have not said you know each other." followed by a checkbox "You know Ted" which is currently unchecked. Below that, it says "You and Ted know no people found in common". Under the heading "Basic Information", there is a "Look up" section with a button labeled "Qdos reverse links".

On the right side of the pane, there is a section titled "Edit your Respect My Privacy restrictions!". It contains several checkboxes: "no-commercial" (checked), "no-depiction" (unchecked), "no-employment" (unchecked), "no-financial" (checked), and "no-medical" (checked). Below these is a "Save my Respect My Privacy Restrictions" button. To the right of this section is a "Respect My Privacy" icon with three sub-icons: "can", "in", and "not". Below the icon are two buttons: "Edit Respect My Privacy Restrictions" and "Forget my Web ID".

Below the RMP section is an "Acquaintances" list with the following entries: Alex Chao, Brian Lee, Faye Yu, Greg Soltis, Helen Drislane, Jason Yow, and Ji Won Yang.

# Privacy/License Highlighter in Tabulator

The screenshot shows a web browser window with the URL `http://dig.csail.mit.edu/2008/02/rmp/bob-foaf.n3`. The page content is structured as follows:

- Restricts**
  - ▶ No-Medical
- Type**
  - ▶ RDFDocument
- Mentions**
  - ▶ Person
  - ▶ PersonalProfileDocument
- Maker**
  - ▶ **Bob Biggins**
    - Type ▶ Person
    - Family\_name Biggins
    - Given name Bob
    - Image
    - Knows
      - ▶ Alice Someone
      - ▶ Carol Anyone
    - Name Bob Biggins
    - Is developer of
      - ▶ **Social Web Privacy**
        - License ▶ CC:Attribution-NonCommercial-NoDerivs
        - Description This project is focussed on helping users conform to existing privacy policies data. The project is implemented across several platforms - Facebook, Open
        - Developer
          - ▶ Alice Someone
          - ▶ Bob Biggins
- ▶ **Carol Anyone**
  - Type ▶ Person
  - Family\_name Anyone
  - Given name Carol
  - Image
  - Knows ▶ Alice Someone
  - Name Carol Anyone
  - Is developer of ▶ Social Web Privacy
  - Is knows of ▶ Bob Biggins
  - Made
    - ▶ `http://dig.csail.mit.edu/2008/02/rmp/carol-foaf`
    - ▶ `http://dig.csail.mit.edu/2008/02/rmp/carol-foaf.n3`

- Helper**
- ▶ `http://csail.mit.edu/~lkagal/foaf#me`
- ▶ `http://dig.csail.mit.edu/2008/02/rmp/tkang-foaf#me`
- Homepage**
- ▶ `http://dig.csail.mit.edu/2009/SocialWebPrivacy/`
- Name** Social Web Privacy
- Programming language** Javascript
- Type** ▶ Project
- Is knows of** ▶ Alice Someone
- Made** ▶ `http://dig.csail.mit.edu/2008/02/rmp/bob-foaf`



# References

---

- Policy Aware Content Reuse on the Web: <http://dig.csail.mit.edu/2009/Papers/ISWC/policy-aware-reuse/cc.pdf>
- Respect My Privacy: <http://dig.csail.mit.edu/2010/Papers/Privacy2010/tkang-rmp/paper.pdf>
- Information Accountability: <http://dspace.mit.edu/bitstream/handle/1721.1/37600/MIT-CSAIL-TR-2007-034.pdf>



# Questions?

---

[oshani@mit.edu](mailto:oshani@mit.edu)