The Problem

Reusing content saves resources and fosters creativity. However, reusing a particular piece of content without honoring the license expressed with it may violate the original content creator’s rights. There are several reasons this situation might happen. The person reusing the content may be: 
- too lazy to check for the licenses hidden in the XHTML
- weary of the multi-step operations required to embed the license metadata
- ignorant as to what each of the licenses mean

At the same time, the original content creator would also be interested in knowing whether someone has violated his or her license terms.

How much of a problem is this?

Flickr has over 100 million Creative Commons Licensed images. Given a sample of web pages which embed such images, how many of these are properly attributed as specified in their licenses?

A simple experiment was conducted to get an assessment on this, and the results are as follows:

- **Sample 1**
  - Properly attributed images = 28
  - Misattributed images = 333
  - Misattribution = 78 %
- **Sample 2**
  - Properly attributed images = 8
  - Misattributed images = 194
  - Misattribution = 80 %
- **Sample 3**
  - Properly attributed images = 6
  - Misattributed images = 439
  - Misattribution = 94 %

Build Policy Aware Systems, such as:
- Validators to tell users what information is missing or inaccurate
- Seamlessly integrate metadata by detecting and assisting in embedding the licenses
- Notify users if their content is used in an inappropriate manner

Background

Policies are pervasive in web applications as they play a crucial role in enhancing security, privacy and usability of services offered on the Web. Use of Creative Commons licenses is the widely accepted method of expressing rights of the original content creators when it comes to digital multimedia content on the Web.

How can you Extract License Metadata?

1. Through APIs which expose the licenses, E.g. Flickr
2. Through RDFa (Resource Description Framework in Attributes)

A simple scenario which illustrates a rights violation of a content creator:

*Illustration of a license incompliant content reuse*

The Solution

Enable transfer of content between Web applications with minimal effort in a policy aware manner, i.e. when content is copied, license metadata is also copied and pasted appropriately in the target application.

**Goal**

Enable transfer of content between Web applications with minimal effort in a policy aware manner, i.e. when content is copied, license metadata is also copied and pasted appropriately in the target application.

**Components**

- **RDFa Extractor**: Extracts all the semantic information in the form of RDF attributes embedded in the HTML page the user browses.
- **UI Enhancer**: Adds visual cues to the page for easy identification of images that can be copied based on the user’s intended use.
- **RDFa License Store**: Indexes the License data of images in a given browser session.
- **Attribution XHTML Constructor**: Creates the attribute XHTML snippet as stated in the CC specification upon a copy instruction. Then it places this snippet in the system clipboard.

**Future Work**

- Assess the level of violations with regards to other types of licenses such as ‘no commercial use’, ‘share alike’ and ‘no derivatives’
- Assess the level of license violations on other types of media
- Extend to licenses embedded in free-floating content
- Explore new and efficient ways of license violations detection
- Improve the User Interfaces of the CC license violations validator and the Semantic Clipboard