

# THE SCALE AND WAYS OF USING THE CREATIVE COMMONS LICENSE IN POLAND. AN ANALYSIS OF LICENSE'S USAGE ON WEB PAGES.

## Executive summary.

dr Alek Tarkowski, Ewa Majdecka

In 2013, we conducted an experimental indexing of Polish Internet resources. Our primary aim was to measure the usage scale of Creative Commons licenses. The amount of content under CC licenses is a basic measure of the development of the free culture movement. As a result, we also obtained a representative sample of the pages made available under CC licenses or those that contain content on this license. It has allowed for an analysis of the methods and accuracy of copyright information regarding CC licenses.

Despite the passage of time from the moment of the study's implementation, we are convinced that the study data provides useful knowledge on the use of CC licenses in Poland. Its main advantage is it being based on an empirical analysis of the use of CC licenses, which was conducted on a large sample of web pages.

The analysis was conducted on the basis of a list of web pages containing the phrase "Creative Commons," indexed by a web crawler in October-November of 2013. 19.6 million web pages (single-HTML documents) were indexed under 735 thousand Internet domains, which were identified with using the Polish language.

### THE USAGE SCALE OF CC LICENSES ON THE POLISH INTERNET

**The main result of our study was data regarding the usage scale of CC licenses on the Polish Internet.** Limitations occurred due to the difficulty in defining a complete collection of Polish web pages - which makes it impossible to provide the actual number of Polish web pages (or the general content) under a CC license. However, with a very large sample of 19.6 million web pages, we can - with a high probability - estimate the share of CC licensed sites. 1.25% of the web pages contain the phrase "Creative Commons," and 0.68% of the web pages contain links to canonical pages with the text of one of the CC licenses. Similarly, 0.48% of domains contain the phrase, and 0.22% of domains contain a link to the license. (The difference in the results is related to the fact that sometimes one



domain contains more than one page under a CC license). **We assume, rounding the result, that about 1% of all online content is under one of the CC licenses.**

## **Web pages under one of the Creative Commons licenses**

The basic requirement for licensing content under a CC license is to provide licensing information. For that reason, we checked the way of presenting this information on a page. **70% of domains using a CC license place information regarding a selected Creative Commons license in the footer of the page.**

60% of domains present the license by providing the name or symbol of a specific license and a link to a page with the full license content. It is a good result, suggesting large-scale understanding in this group of how to utilize CC licensing. However, **almost 20% of sites committed a fundamental error, by not indicating the specific license in use.** (It should be noted that we were unable to measure the scale of use of CC licensed content without any attribution or licensing information - as it did not appear in our sample).

**The most commonly used Creative Commons license, in almost 1/4 of the cases, was BY-SA (Attribution - Share-alike), also used by Wikipedia.** Interestingly, the four most popular Creative Commons licenses on the Polish Internet are the least and most restrictive. It is also worth noting that **almost 50% of the web pages are available on one of the two free licenses (CC BY or CC BY-SA).**

## **Web pages with content shared under Creative Commons licenses**

**The vast majority of the content shared under CC licenses consists of images (49%).** The next categories are shared much less frequently. Ebooks and text files (usually in a PDF format) constitute 12%, and 11% are texts published directly on the web page, e.g. posts on blogs.

One of the reasons for this is the role of major platforms, which share images under CC licenses, which are then used on other web pages. Flickr or Wikimedia Commons make it possible to license content with a selected (Flickr) or specific (Wikimedia Commons) CC license.

▫ **of the analyzed content has not been produced by the authors of particular web pages, but downloaded from other sources.** The most common source was Flickr. Even if information about a Creative Commons license was missing under an image, it was present in the metadata of the images and in this way, it was indexed by a web crawler. **1/3 of the content found under Creative Commons licenses was created by the authors of the analyzed web pages.**



Similarly to web pages that were in their entirety shared under CC, content by other authors, when used on a webpage, usually contains both information about the licensing conditions and a link to the web page with the full license content. The scale is, however, smaller - just over 40%, compared to 60%.

## The systematic or incidental use of a Creative Commons license

The 243,544 pages available under CC, or which contain content on this license, are unevenly distributed between domains. 916 domains had just one link related to Creative Commons, 825 domains had from 2 to 49 links on a page. This is one of the most significant conclusions of this study. **In half the cases, using a Creative Commons license turns out to be incidental or even a one-time occurrence.** It is important to take into account that quite often the licensed content consists of images and graphics, and these are largely drawn from the largest services with images and graphics (such as Flickr and Wikimedia Commons). Content from these portals have information about CC in the metadata, and upon copying to one's own website, the signature as well as a link to the Creative Commons page is automatically downloaded. They can be indexed by web crawlers, but users may not even know that they are using a work (image, graphic, etc) under a CC license.

## CONCLUSIONS AND RECOMMENDATIONS

Our study confirmed the often accepted scale of availability of content under Creative Commons licenses - about 1% of all content. This relatively low percentage translates into a great amount of content available under free licenses.

More significantly, CC licensing of content is a result of the informed decisions of people, who decide to make content available differently than under the automatically operating rule of "all rights reserved." Another significant finding is also that half of the content is available under one of two free licenses: Attribution CC or CC Attribution - Share Alike.

Crucial conclusions from the analyses concern ways of making content available. 60% of all content is properly marked - including the full name of the license along with its indicated conditions, and a link to the web page with the license text. This is a positive result, attesting to the fact that the persons deciding to make content available under CC licenses have a relatively high knowledge of copyrights laws. At the same time, 40% of cases include situations where the content is - in various ways - incorrectly marked. Further education and the raising of awareness regarding the correct use of CC licenses can play a considerable role in rectifying the situations in which the content is incorrectly marked.



When using other people's content, available under CC licenses, the level of correct indications drops to 40%. This is an important result, indicating that people who do not themselves share content are much less likely to understand the principles of applying this kind of license. This is connected to the fact that use of content under CC licenses is often a one-time or incidental event.

Finally, images hold a dominant role among content made available and used under CC licenses. Online platforms, such as Wikimedia Commons and Flickr, play a crucial role for making images available for further use. Their distinct influence on the use of images under CC licenses by Internet users indicates the major role that the introduction of CC licensing on major content platforms can play.

Further studies should go beyond quantitative analysis of the use of CC licenses. We should aim at answering questions about the motivation of people who share and use content, also regarding the chosen license.



This text is available under a Creative Commons Attribution 3.0 Poland license (<https://creativecommons.org/licenses/by/3.0/pl/>). Some rights reserved.

Contact: dr Alek Tarkowski: [atarkowski@centrumcyfrowe.pl](mailto:atarkowski@centrumcyfrowe.pl), Ewa Majdecka: [emajdecka@centrumcyfrowe.pl](mailto:emajdecka@centrumcyfrowe.pl)