Commons Copyright Templates
October 2016
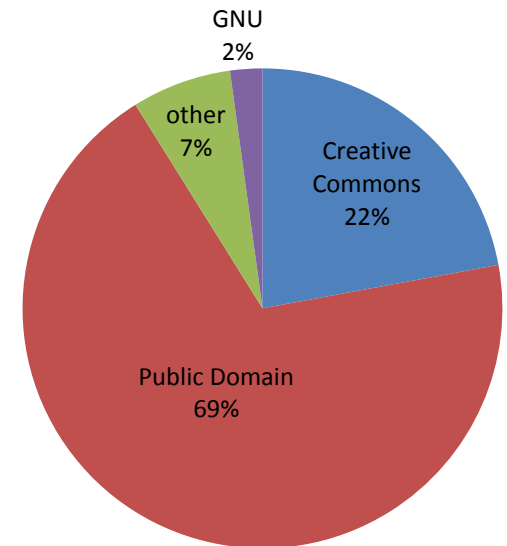
By Jarek Tuszynski

(user:jarekt)
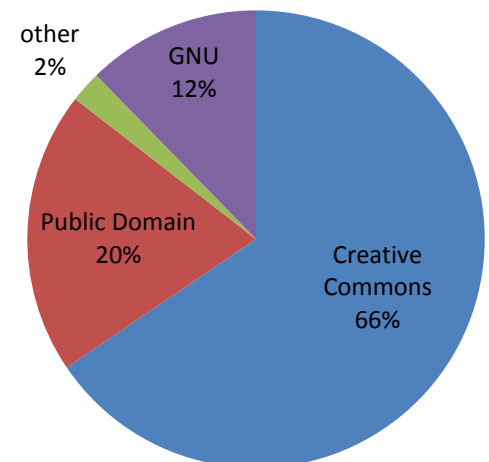
Overview

Copyright templates by type

- 4167 templates that are or contain a copyright templates (aka, *license templates*)
- 971 unique or "primary" license templates
  - Meant to be used in files (not to build other templates)
  - If used then file satisfies requirement of having a license
  - Do not transclude other primary license templates
  - In [[Category:Primary license tags (flat list)]]
- Why so many?
  - 69% of templates are *public domain* (PD) templates, many for individual countries, but only 20% of files use them
  - 22% of templates are *Creative Commons* (CC) templates which are used by 66% of files. Many CC templates are for jurisdiction ports for different countries.
- All templates are PD or free licensed .
  - We no longer have any non-free licensed files, thanks to Wikimedia who in October 2014 relicensed all their logos from non-free license to CC-BY-SA
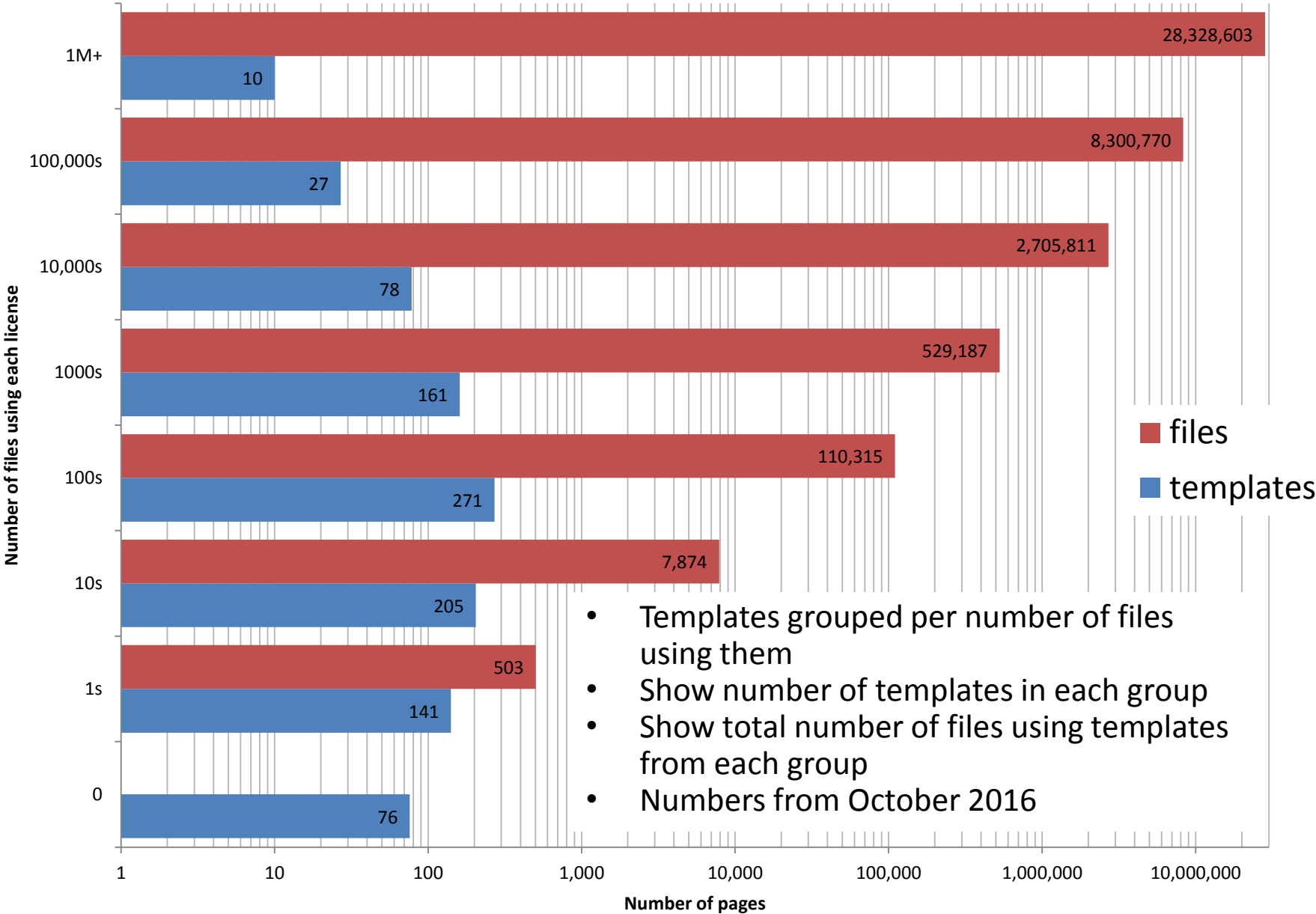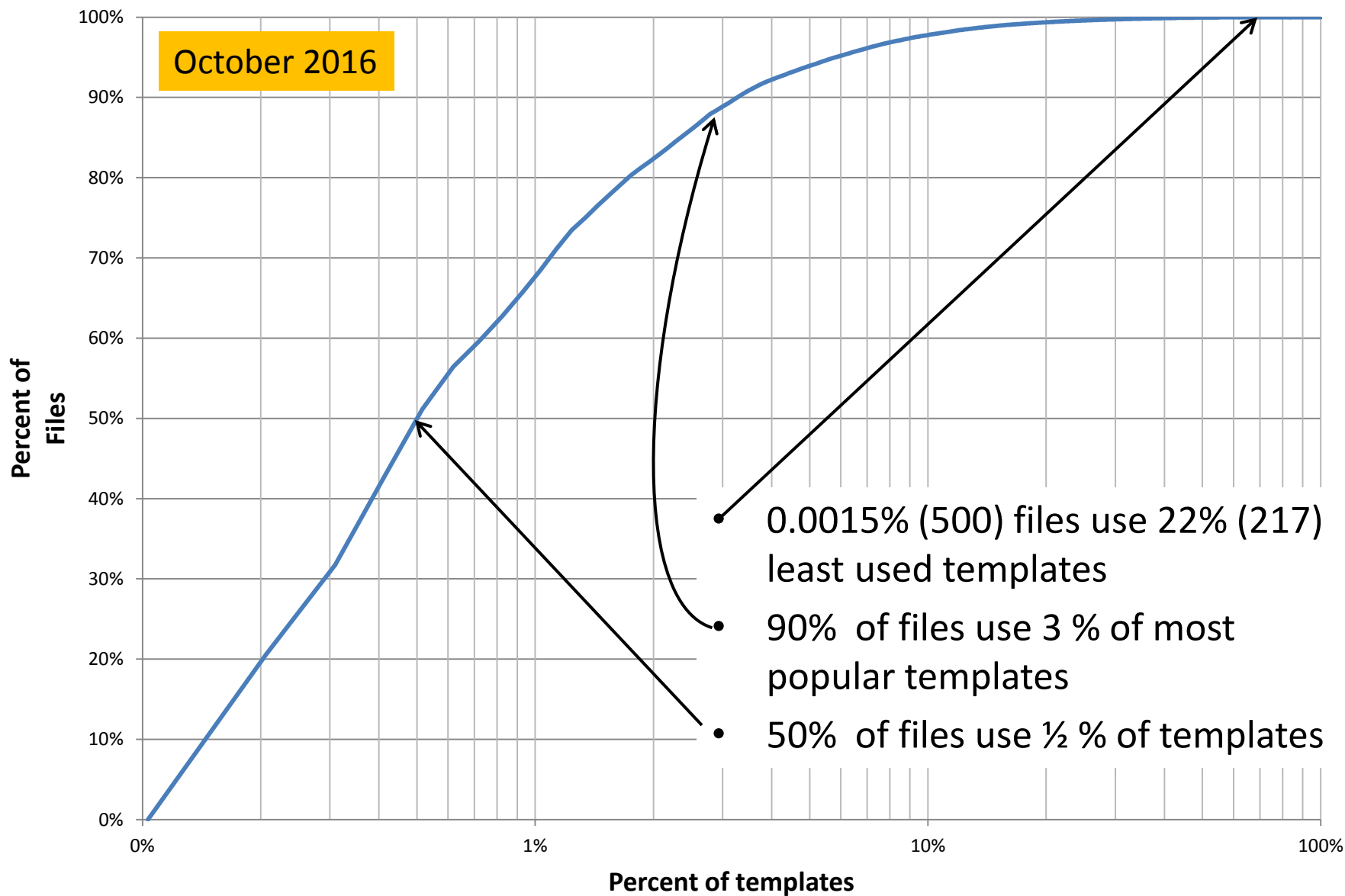
Copyright templates by type

GNU 2%
other 7%
Creative Commons 22%
Public Domain 69%

Files using different types of copyright templates

other 2%
GNU 12%
Public Domain 20%
Creative Commons 66%

Statistics per number of files using each template

Number of files using each license (y-axis)

Category	files	templates
1M+	28,328,603	10
100,000s	8,300,770	27
10,000s	2,705,811	78
1000s	529,187	161
100s	110,315	271
10s	7,874	205
1s	503	141
0		76

Number of pages (x-axis)

- Templates grouped per number of files using them
- Show number of templates in each group
- Show total number of files using templates from each group
- Numbers from October 2016

Number of files using most popular license templates

October 2016

Percent of Files (y-axis): 0% to 100%
Percent of templates (x-axis): 0%, 1%, 10%, 100%

0.0015% (500) files use 22% (217) least used templates

90% of files use 3 % of most popular templates
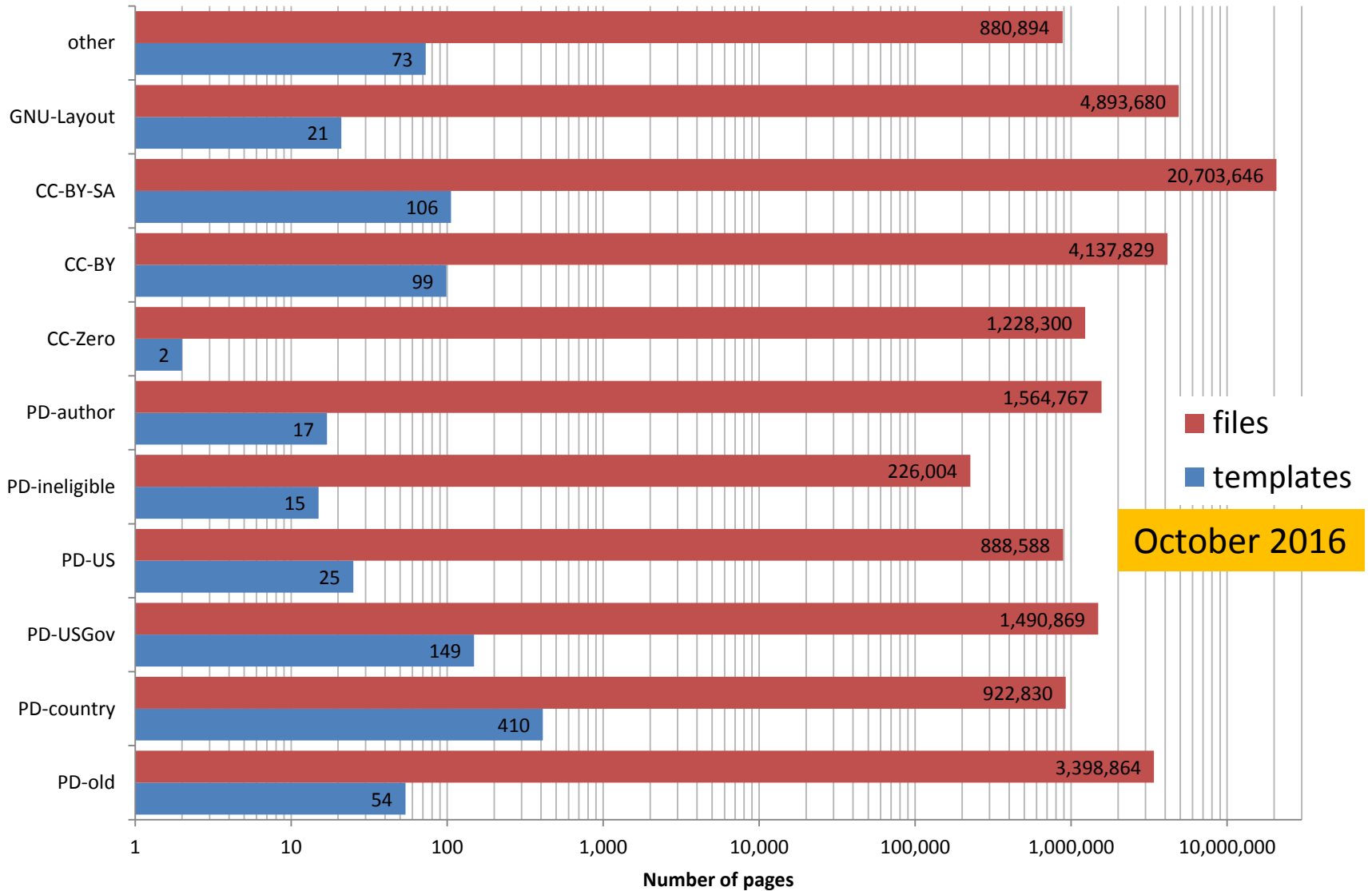
50% of files use ½ % of templates

Public Domain Templates

- Largest group of copyright templates are Public Domain (PD) templates, with many subgroups, each named after most generic copyright template:
  - **PD-old** – copyrights expired due to age (includes "no known restrictions" templates for various institutions)
  - **PD-country** – templates for files in PD based on laws of specific countries (except US)
  - **PD-US** - templates for files in PD based on US law (except US government works)
  - **PD-USGov** - templates for files in PD based on laws for works of US government employees
  - **PD-ineligible** - templates for files in PD because the content is not eligible for copyrights
  - **PD-author** – templates for cases when copyright owners released the files to PD (except for CC-zero)

License templates for copyrighted works

- 31% of templates and 80% of files are copyrighted:
  - Creative Commons licenses: CC-by-sa, CC-by and CC-zero (CC-zero is very close to PD-author)
  - GNU licenses: GFDL, GPL, LGPL, etc.
  - Other:
    - Attribution licenses
    - Country specific non-PD licenses for copyrighted material
    - {{Beerware}}, {{CeCILL}}, etc.

Number of files and templates for each group of licenses
(grouping by families of templates)

License group	files	templates
other	880,894	73
GNU-Layout	4,893,680	21
CC-BY-SA	20,703,646	106
CC-BY	4,137,829	99
CC-Zero	1,228,300	2
PD-author	1,564,767	17
PD-ineligible	226,004	15
PD-US	888,588	25
PD-USGov	1,490,869	149
PD-country	922,830	410
PD-old	3,398,864	54

files
templates

October 2016

Number of pages

Roles of copyright templates

- All files are required to use (transclude) an "official" copyright template from [[Category:Primary license tags (flat list)]].
- License templates serve several purposes:
  - **Translations** – Presenting the text in the language of the user. Challenge: updating translations after English version changes.
  - **Verification** – Presenting text that is clear, concise and correct.
    - In case of PD templates we verify that there is a law which places some group of files that meet some conditions in PD
    - In case of copyrighted images, we verify hat the license meets requirements of a free license.
  - **Machine readability** – Allow tools to automatically recognize that there is a license and what type of license it is
- License templates with more transclusions, have more visibility and are more likely to fulfill those purposes.

License enforcement

- Q: How is license use enforced? A:Through two parallels systems:
  - In 2011 we created empty {{License template tag}} template which is included in several layout templates used by all the license templates. Query for files not using {{License template tag}} finds all the files without a license.
  - In 2013 MediaWiki software began using special machine-readable tags to encode license info. The tags are added by license templates, and are used by many tools including Media Viewer. Files without them are automatically added to [[Category:Files with no machine-readable license]].
- Every day a bot adds files not using {{License template tag}} to
  - [[Category:New uploads without a license]]
  - [[Category:Media without a license: needs history check]]
- Having a license does not guarantee that license is appropriate, for this we still need human verification

Changes to the primary copyright template list

- Effort to change PD templates from generic to more specific
  - {{PD}} →{{PD-old}} → {{PD-old-100-1923}}
- Some unused templates are retired or deleted (76 remains)
- Some templates are rewritten as *secondary* licenses relying on more generic *primary* licenses. This process helps with maintenance of rarely used templates.
- Laws change and some PD templates become invalid
  - Extending copyright term retroactively, like Russian law from 2006
  - German stamps are no longer PD after court case
- More small niche PD templates based on so far unused laws, like US WWII Alien_Property laws.
- Ongoing discussions about borderline cases
  - Can uploaders create their own *homebrewed* license templates? For example 2 users copied the text of CC-by-sa but added resolution restriction which allows them to have low-resolution images on Commons under one license and high-resolution images somewhere else under CC-BY-NC license.
  - Many people use GFDL-1.2 (valid but unusable for images) multi-licensed with CC-BY-NC (un-free license). How do we limit that?
  - What to do with PD due to "no known restrictions" statements used by many institutions?

References

- Numbers based on
  - https://quarry.wmflabs.org/query/5441
  - https://quarry.wmflabs.org/query/5201
- Groupings based on
  - Use of {{PD-Layout}}, {{GNU-Layout}}, {{CC-Layout}}, {{Copyrighted-Layout}} and {{Other License-Layout}} templates (see https://commons.wikimedia.org/wiki/Template:Style_formatting_license_template/doc)
  - My reading of the text of the copyright template
  - Can be imperfect