CREATIVE COMMONS AND LANGUAGE RESOURCES: GENERAL ISSUES AND WHAT'S NEW IN CC 4.0

Pawel Kamocki & Erik Ketzan

Creative Commons ("CC") licenses are important and powerful tools in the creation and sharing of language resources. The first part of this paper discusses some general issues and common misconceptions regarding CC and language resources. The second part highlights the changes in Creative Commons Version 4.0 ("CC 4.0", released November 25, 2013) most relevant to language resources. This paper assumes a basic familiarity with Creative Commons licenses. For an introduction to them, see creativecommons.org

Version 1.1, August 2014: Sections 8 and 12 of this paper were expanded and modified for clarity. Correction: Section 4 in the initial version of this document stated inaccurately that works "may only" be re-licensed under certain circumstances, when in fact it "only makes sense" to re-license them under those conditions.

Please note: This publication may reflect current thinking or ongoing discussion within CLARIN, but it does not necessarily reflect the position of CLARIN ERIC, unless indicated otherwise.

Kamocki and Ketzan are Research Associates at Institut für Deutsche Sprache (Mannheim, Germany). Contact: kamocki@ids-mannheim.de, ketzan@ids-mannheim.de

CC LICENSES AND LANGUAGE RESOURCES

1) CC licenses are not intended for software

2) Only some of the CC licenses meet the definition of "open," "open content," or "open license", and this has a different meaning from "open access"

3) There is community-wide confusion regarding what types of use are "non-commercial" for the purposes of CC-NC

4) Changing / updating the CC licenses of things already licensed under a CC

5) Courts and institutions across Europe have shown widespread support for the general validity of CC licenses

6) A word of caution about CC0 ("CC-zero")

7) *CCPlus – a little known tool that may be useful for certain resources*

WHAT'S NEW IN CC 4.0

8) CC 4.0 licenses are designed to be "universal"; there are no ported (national) versions of the CC 4.0 licenses

9) CC 4.0, unlike its predecessors, assigns not only copyright, but also the *sui generis* database right

10) Now that it addresses databases, CC 4.0 is better than other open licenses for language resources

11) Moral rights are (still) not licensed under CC 4.0

12) CC 4.0 includes a slight change to attribution under the BY module, designed to better reflect accepted practices

CC LICENSES AND LANGUAGE RESOURCES: OVERVIEW

CC LICENSE	GENERAL DEFINITION[1]	SPECIAL REMARKS RE: LANGUAGE RESOURCES
Attribution **CC BY**	This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.	• CC 4.0 specifically allows **links to separate pages** to satisfy the attribution requirement. • For **requirements and best practices** of attribution, see section 13 below. • Language resource metadata or a linked file should contain every element required. • Although the BY module sounds ideal for language resources, in practice, **implementing the attribution requirements presents serious challenges**. • By far the license **most commonly used in Open Access** (see section 2 below)
Attribution-ShareAlike **CC BY-SA**	Lets others remix, tweak, and build upon a work even for commercial purposes, as long as it is attributed and license their new creations under the identical terms. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia.	• See remarks for CC BY • Language resources using these works must be released under the same CC license or a "Creative Commons Compatible License". As of May 2014, Creative Commons has not approved any compatible licenses, but plans to at some future date.[2]
Attribution-NoDerivs **BY-ND**	Allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with attribution.	• See remarks for CC BY • In general, **not a good choice for language resources**, which annotate, re-order, modify, and transform data. • The **definition** of what constitutes a derivative work **varies by country**, and it is unclear what types of language resources constitute derivatives • Arguably, lexical work, corpus compilation, and concordances, for example, are derivative works in many jurisdictions. • Anecdotally, researchers working with language resources have been hesitant to annotate BY-ND works.

---

[1] Definitions mostly from the official descriptions at Creative Commons, "About the Licenses", available at: http://creativecommons.org/licenses/ (last visited May 20, 2014, as were all links mentioned in the footnotes here).

[2] Creative Commons plans to list Creative Commons Compatible Licenses at "Compatible Licenses", available at: http://creativecommons.org/compatiblelicenses

CC LICENSE	GENERAL DEFINITION	SPECIAL REMARKS RE: LANGUAGE RESOURCES
Attribution-NonCommercial **CC BY-NC**	Lets others remix, tweak, and build upon a work non-commercially, and although their new works must also give attribution and be non-commercial, they don't have to license their derivative works on the same terms.	• See remarks for CC BY • **There is community-wide confusion over what "non-commercial" means** (see section 3, below) • The CC community, Wikipedia, and Wikimedia-backed commentators all suggest that **the SA module is superior to NC for many purposes** (see section 3 below).
Attribution-NonCommercial-ShareAlike **CC BY-NC-SA**	Lets others remix, tweak, and build upon a work non-commercially, as long as they give attribution and license their new creations under the identical terms.	• See remarks for CC BY-NC • See remarks for CC BY-SA
Attribution-NonCommercial-NoDerivs **CC BY-NC-ND**	The most restrictive of the six main CC licenses, only allowing others to download works and share them with others as long as they give attribution, but the work cannot be changed in any way or used commercially.	• See remarks for CC BY-NC • See remarks for CC BY-ND

1) CC licenses are not intended for software

CC explicitly recommends against using most CC licenses for software, as CC licenses do not mention source or object code, and are not compatible with the GPL (GNU General Public License).[3] Creative Commons does suggest the use of CC0 (CC-Zero) may be appropriate for software (in part because it is GPL compatible).[4]

GPL and other open software licenses are more appropriate for software.

As of CC 4.0, the licenses are appropriate for databases (see section 9 below).

2) Only some of the CC licenses meet the definition of "open," "open content," or "open license", and this has a different meaning from "open access"

Researchers dealing with language resources will sometimes come across requirements for "open content" or "open licenses" in the course of their work. Sometimes funding documents, grant applications, and other project requirements will stipulate that certain content must be published or distributed in an "open" manner or under an "open license". Be aware that only some CC licenses meet this definition.

Under the widely-recognized "Open Definition"[5] of the Open Knowledge Foundation, only three CC licenses qualify as "open": CC-BY (and its 4.0 iteration), CC-BY-SA (and its 4.0 iteration), and CC0 ("CC-Zero").[6]

"Open access" means something more specific than general open content. Open access (OA) refers particularly to scholarly articles and publications. Although OA has long been important across all research fields, it is especially relevant now because, as of 2014, member states are beginning to implement laws reforming OA following EU guidance in 2012 (2012/28/EU).

Again, only certain CC licenses are fully consistent with "open access" publication. "Open access" has been defined by community across various fora,[7] the most recent of which define an open access where users must be able to "copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship."[8]

---

[3] Creative Commons, "Frequently Asked Questions", available at:
http://wiki.creativecommons.org/Frequently_Asked_Questions#Can_I_use_a_Creative_Commons_license_f
or_software.3F

[4] Creative Commons, "CC0 FAQ", available at: http://wiki.creativecommons.org/CC0_FAQ ("CC0 is suitable for dedicating your copyright and related rights in computer software to the public domain, to the fullest extent possible under law.").

[5] Open Knowledge Foundation, "Open Definition, Version 1.1", available at: http://opendefinition.org/od/

[6] Open Knowledge Foundation, "Conformant Licenses", available at: http://opendefinition.org/licenses/

[7] Most importantly, the the Budapest Open Access Initiative in February 2002, the Bethesda Statement on Open Access Publishing in June 2003, and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in October 2003. See Peter Suber, Open Access, (The MIT Press Essential Knowledge Series ed.). Cambridge, Mass.: MIT Press, at pp. 7-8.

[8] See Bethesda Statement on Open Access Publishing, 2003, available at:
http://dash.harvard.edu/handle/1/4725199 and Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003, available at: http://openaccess.mpg.de/286432/Berlin-Declaration

The CC-BY is by far the most widely used and recommended in the open access community.[9]

3) There is community-wide confusion regarding what types of use are "non-commercial" for the purposes of CC-NC

There is widespread confusion regarding the "non-commercial" requirement of CC-NC.[10] Consequently, the NC module is less than ideal for many language resource uses.

John Weitzmann, Legal Project Lead of Creative Commons Germany, wrote in 2012:

> ... the restriction NC – NONCOMMERCIAL USE ONLY – is very popular with the authors... However, deciding on a license that does not allow commercial use has extensive consequences: Many possibilities of use, like the inclusion in knowledge communities and archives, the Wikipedia, local newspapers, publications, compilations and mashups, are in fact excluded and thus require additional approval – and that despite the fact that these inclusions are often wanted by the authors.[11]

This statement would probably come as a surprise to many researchers with only superficial knowledge of CC, especially in the case of Wikipedia (which is a project of the Wikimedia Foundation, an American non-profit and charitable organization). "If Wikipedia is not "non-commercial", what is?" is not an unreasonable question.[12] In fact, Wikipedia generally uses content licensed CC-BY-SA, in part because it bypasses gray zone questions of what is commercial and what is not.

Discussing the contours of non-commercial activity, Dr. Paul Klimpel writes that a course at a public university that charges course fees creates a commercial situation, and a blog that displays advertisements is not unambiguously non-commercial.[13] As a general definition, Klimpel suggests that commercial use should be assumed "for anyone who is not fully financed by public funds or private donations."[14] He notes, however, that, "there are very few not-for-profit institutions these days that have adequate

---

[9] Creative Commons, "Creative Commons and Open Access", available at:
http://wiki.creativecommons.org/Creative_Commons_and_Open_Access

[10] See, e.g.: University of California libraries Creative Commons License Working Group Final Report, 30 April 2013, at p. 2, available at:
http://libraries.universityofcalifornia.edu/groups/files/sag1/docs/Creative_Commons_License_WG_Report_Final.pdf ("There is some confusion and disagreement, even among those who are extremely familiar with Creative Commons licenses, about the meaning of the NonCommercial restriction."); comments by Eric Raymond, founder of the Open Source Initiative and prominent figure in the open source movement, under Creative Commons, "Ongoing Discussions: NonCommercial and NoDerivatives", 29 August 2012, available at: http://creativecommons.org/weblog/entry/33874#comment-372240 ("The NC option in Creative Commons has always been a bad idea and should be removed... [because of] the fact that there is no bright-line legal test for "commercial activity""); Gordon Haff, "Does the Noncommercial Creative Commons license make sense?", 27 November 2007, available at: http://www.cnet.com/news/does-the-noncommercial-creative-commons-license-make-sense/ ("[NC] seems a dangerously ill-defined question in an environment where individuals have so many opportunities to micro-commercialize").

[11] Paul Klimpel, "Consequences, risks and side-effects of the license module "non-commercial use only – NC", available at: http://openglam.org/files/2013/01/iRights_CC-NC_Guide_English.pdf

[12] Klimpel describes his reasoning as to why NC content cannot be used in Wikipedia in Part 8 of Ibid.

[13] Id. at p. 10.

[14] Id. at p. 11.

funding and do not depend on additional revenues – which again places them in the gray area of our attempted distinction. To completely avoid the NC module and its restrictions will avoid these uncertainties."[15] Again, Klimpel suggests the SA module as a superior choice for both authors and people relying on CC licenses.

In March 2014, the confusion regarding "non-commercial" played out in a German regional court, the Landgericht Köln.[16] There, the Court held that placing a CC-licensed photo on the website of Deutschlandradio, part of the German public broadcaster, was a commercial use, even though the public radio's website contained no advertising, charged no fee, and had no sponsorship. Although this decision may get reversed on appeal, it further demonstrates the confusion regarding what "non-commercial" means.

In summary, the CC community, Wikipedia, and Wikimedia-backed commentators all suggest that the SA module is superior to the NC module for many purposes that sound "non-commercial" but may fall into a legal gray zone. Purposes such as research with language resources.

4) Changing / updating the CC licenses of things already licensed under a CC

Creative Commons licenses, by their terms, are irrevocable.[17] In other words, if you released a resource in the past under a CC license, that license will remain valid. If for some reason *you* wish to re-release it under a new license, this generally only makes sense if:

- the **new license is more liberal** than the old one
    - o Example: something licensed under CC-BY-SA 3.0 could be re-licensed under CC-BY 3.0/4.0.
    - o Explanation: it does not make sense to re-license something under a more restrictive license, because the old one (as originally distributed) will remain in effect.

- or, through **updates of the CC licenses**
    - o Example: content previously licensed under a CC 3.0 license may be re-licensed under that CC license's 4.0 equivalent.

5) Courts and institutions across Europe have shown widespread support for the general validity of CC licenses

CC 4.0 is naturally too new to have been tested in courts, but the history of CC in Europe shows consistent, strong recognition from courts and other institutions regarding the general enforceability of CC licenses. Courts in Germany, Belgium,

---

[15] Ibid.

[16] LG Köln v. 05.03.2014, Az. 28 O 232/18, available at: http://openjur.de/u/686021.html

[17] See, e.g., Creative Commons, "Attribution 4.0 International", available at: http://creativecommons.org/licenses/by/4.0/legalcode ("Considerations for licensors... Our licenses are irrevocable"). For an analysis of the problems of irrevocability in open licenses under United States law, see Timothy K. Armstrong, "Shrinking the Commons: Termination of Copyright Licenses and Transfers for the Benefit of the Public", Harv. J. on Legis. 47, 359. (2010).

Netherlands, Spain, and numerous other jurisdictions have either explicitly or implicitly upheld the general validity of CC licenses.[18]

6) A word of caution about CC0 ("CC-zero")

CC0 ("CC-Zero") was released in 2009, and is a "no copyright reserved" tool intended to allow rightsholders to relinquish all copyright in a work and dedicate it to the public domain.[19] CC-Zero-designated content is ideal in theory for language resources because it would allow unfettered use of such material. A number of major European institutions make use of CC-Zero, including Dutch National Government, CERN, the British Library, National Library of Spain, Swedish National Library, plus many others.[20]

There are two potential problems with using CC-Zero for language resources.

First, the enforceability of CC-Zero is uncertain in European jurisdictions.

By design, CC-Zero allows an author to waive their copyrights, but many European jurisdictions do not allow an author to do this (originally, this was designed for his/her protection). Many Member States do not allow an author to waive his moral rights (i.e. the rights protecting the work against distortion or guaranteeing the right of authorship attribution), while other do not allow an author to waive *or* transfer moral rights (this is the case in e.g. France and in Germany, while in the United Kingdom, on the other hand, moral rights can be waived).

CC-Zero compensates for this by its "fallback provision", which states that the waiver is not effective for any reason (such as the fact that moral rights cannot be waived in that jurisdiction), then CC-Zero acts as a license from the affirmer granting the public an unconditional, irrevocable, non exclusive, royalty free license to use the work for any purpose.[21]

Europeana, the European Union-funded digital archive, has taken the position that CC-Zero may be used in all European jurisdictions.[22] A legal memorandum commissioned by Europeana, by German attorney Dr. Till Kreutzer, concluded that although the CC-Zero waiver itself is not legally valid under German law insofar it relates to the author's rights, the "fallback provision" in section 3 of CC-Zero is valid under German contract and copyright law, and largely accomplishes the goals of any CC-Zero waiver.[23]

Nevertheless, doubts as to CC-Zero's enforceability have been raised in many venues. During Europeana's 2011 call for feedback on its Europeana Data Exchange Agreement (DEA), organizations from Sweden and Poland questioned CC-Zero's

---

[18] Creative Commons, "Case Law", available at: http://wiki.creativecommons.org/Case_Law

[19] Creative Commons, "CC0", available at: http://wiki.creativecommons.org/CC0

[20] Creative Commons, "CC0 use for data", available at: http://wiki.creativecommons.org/CC0_use_for_data

[21] Creative Commons, "CC0 FAQ", available at: http://wiki.creativecommons.org/CC0_FAQ

[22] Europeana, "Analysis of the feedback with regard to the Europeana Data Exchange Agreement (draft as of 05.05.2011)", available at: http://pro.europeana.eu/c/document_library/get_file?uuid=a48beaef-444b-4f23-987a-90c3da23ea7b&groupId=10602

[23] Till Kreutzer, "Validity of the Creative Commons Zero 1.0 Universal Public Domain Dedication and its usability for bibliographic metadata from the perspective of German Copyright Law," i.e. Büro für informationsrechtliche Expertise, available at: http://pro.europeana.eu/c/document_library/get_file?uuid=29552022-0c9f-4b19-b6f3-84aef2c3d1de&groupId=10602

compatibility with their national legislation (the Kreutzer report was commissioned in response to this issue).[24]

Enforceability aside, there are other aspects of CC-Zero that make it less than ideal for language resources. A 2013 deliverable by Linked Heritage, a European Union-funded project to coordinate standards and technologies for Europeana, sharply criticized the use of CC-Zero for metadata, noting that the disclaimer in CC-Zero

> "pushes back the question of the ultimate ownership of data to the final reuser of the data... To reuse CC0-marked data in compliance with database right and existing licenses, thus mitigating the risk of future rights claims, implies difficulty and probably associated costs.
> Therefore taken as a whole without any additional guarantee of up-front rights clearance, **the CC0 waiver has little or no value at all for enabling reuse of metadata, and may in fact make it more difficult**."[25] (emphasis added)

Although the CC-Zero is promising in theory, given this uncertainty, researchers dealing with language resources may wish to favor other CC licenses, both in content they use and content they create, or adopt a "wait and see" approach regarding CC0 until the license becomes more established.

7) CCPlus – a little known tool that may be useful for certain resources

CCPlus ("CC+") is an official Creative Commons license combined with another separate and independent agreement that grants more permissions. In other words, a CC license plus another license. Although it was launched in December 2006, has its own symbol and ways of technical implementation, CC+ remains rather fairly unknown to the general public and to researchers.

We stress that CC+ is not widely used, and for the vast majority of language resources, the primary CC licenses are more practical. But there are special cases of language resources where CC+ could be a good solution:

- CC+ can be used to waive certain requirements of CC license terms or conditions.[26]
- If a researcher finds a CC license that is close, but *not quite* ideal for their resource, CC+ may be a mechanism to still make use of it.
- Release a language resource under CC BY-ND plus a special agreement that researchers involved with language science *may* make derivative works.
- Release a language resource under CC-BY, and if the general standards for attribution are difficult or impossible for such resources, add a CC+ waiver. For instance, "you may satisfy the BY requirement by stating in your NEW project or resource description that MY language resource was used to create it."

For more on implementing a CC+, see http://wiki.creativecommons.org/CCPlus

---

[24] Europeana, supra note 22, at p. 2.

[25] Linked Heritage, "Deliverable D4.3 Specification of legal/licensing environment", available at: www.linkedheritage.eu/getFile.php?id=529

[26] Creative Commons, "Frequently Asked Questions", available at: http://wiki.creativecommons.org/Frequently_Asked_Questions#Can_I_waive_license_terms_or_conditions.3F

W̲HAT'S̲ N̲EW IN̲ CC̲ 4.0

8) CC 4.0 licenses are designed to be "universal"; there are no ported (national) versions of the CC 4.0 licenses

The new CC 4.0 licenses are designed to "work worldwide"; there are currently no ported (i.e. adapted for each nation) versions of the licenses, and few, if any, are planned.[27] There will eventually be official translations (again, not ports) of the 4.0 licenses, and the text of those translations will take precedence over the English text in their jurisdictions.

9) CC 4.0, unlike its predecessors, assigns not only copyright, but also the *sui generis* database right

- In Europe, unlike many jurisdictions (notably the United States), there is a specific legal protection of databases. This protection is called the *sui generis* right.
- Prior to CC 4.0, the CC licenses generally did not mention database rights.
- Some European CC projects attempted to remedy this situation by including databases in version 2.0 – this was true for the ports for France, Germany, Belgium and the Netherlands.[28] This was not seem favorably by the founders of CC, and a compromise was reached such that CC 3.0 would waive the *sui generis* database right in all European CC licenses.[29]
- The need to address the *sui generis* right situation was a primary goal of the 4.0 update.[30]
- This is good for language resources in Europe: the new CC licenses are much more suitable for licensing datasets (such as language corpora) than previous versions.

10) Now that it addresses databases, CC 4.0 is better than other open licenses for language resources

- Before 4.0, the fact that CC licenses did *not* address database rights led other groups to create licenses specifically tailored for databases, notably Open Data Commons (ODC).

---

[27] Creative Commons, "Frequently Asked Questions", available at:
http://wiki.creativecommons.org/Frequently_Asked_Questions#What_if_CC_licenses_have_not_been_ported_to_my_jurisdiction.3F

[28] Lucie Guibault, Licensing Research Data under Open Access Conditions, in D. Beldiman (ed.), *Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance*, Cheltenham, Edward Elgar, at p. 12. Available at:
http://www.ivir.nl/publications/guibault/Open_Research_Data.pdf

[29] Ibid.

[30] Andres Guadamuz, "CC 4.0, an end to porting Creative Commons licenses?", available at
http://www.technollama.co.uk/cc-4-0-an-end-to-porting-creative-commons-licenses

- But now that 4.0 is released, CC is the best choice for research data due to being the most standardized, accepted, and worldwide of these licenses.

11) Moral rights are (still) not licensed under CC 4.0

Most European jurisdictions recognize the concept of moral rights, which include the right of creators to protect the integrity of their work and to receive attribution for their work (among other rights, depending on the jurisdiction). Creative Commons has long held the view that CC licenses do not affect moral rights at all,[31] and CC licenses up to and including CC 4.0 are "intended to minimize the effect of moral rights on otherwise-permitted uses."[32] Although various ported licenses prior to 4.0 contained customized language concerning moral rights in various countries, the new CC 4.0 attempts a global harmonization:

> "Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise."

12) CC 4.0 includes a slight change to attribution under the BY module, designed to better reflect accepted practices

Before 4.0, many people would satisfy the attribution requirement by linking to a separate page that includes the necessary information. Although common practice, there was uncertainty in some communities about whether this was permitted under the licenses. Under 4.0, the CC licenses now explicitly allow this. [33] Additionally, it is no longer mandatory to mention the title of a work to satisfy the attribution requirement (because many works on the Internet only bear a technical title. This is another way Creative Commons has attempted to make the minimum attribution requirements more manageable).

For more on attribution in general, see:

- Creative Commons, "How do I properly attribute material offered under a Creative Commons license?", available at: http://wiki.creativecommons.org/Frequently_Asked_Questions#How_do_I_properly_attribute_material_offered_under_a_Creative_Commons_license.3F
- Creative Commons, "Best practices for attribution", available at: http://wiki.creativecommons.org/Marking/Users

---

[31] Lawrence Lessig, "On the challenge of moral rights", available at http://www.lessig.org/2005/02/on-the-challenge-of-moral-righ/

[32] Creative Commons, "Frequently Asked Questions", available at: http://wiki.creativecommons.org/Frequently_Asked_Questions#How_do_Creative_Commons_licenses_affect_my_moral_rights.2C_if_at_all.3F

[33] Creative Commons, "What's New in 4.0", available at: http://creativecommons.org/Version4