

From Free Culture to Open Data: Technical Requirements for Access and Authorship

Melanie Dulong de Rosnay

► **To cite this version:**

Melanie Dulong de Rosnay. From Free Culture to Open Data: Technical Requirements for Access and Authorship. Danièle Bourcier, Pompeu Casanovas, Melanie Dulong de Rosnay, Catharina Maracke. Intelligent Multimedia - Managing Creative Works in a Digital World, European Press Academic Publishing, pp.47-66, 2010, Series in Legal Information and Communication Technologies. halshs-00671625

HAL Id: halshs-00671625

<https://halshs.archives-ouvertes.fr/halshs-00671625>

Submitted on 17 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From Free Culture to Open Data: Technical Requirements for Access and Authorship

Melanie Dulong de Rosnay*

**Politecnico di Torino*

Abstract. Creative Commons tools makes it easier for users, who are also authors, to share, locate and distribute reusable content, fostering remix and digital creativity, open science and freedom of expression. But reuse could be made even easier by the licensing framework, which does not yet handle the diversity of legal and usage situations pertaining to technical accessibility and reuse modalities of works and data.

This paper will first discuss what additional legal regulation may be required to allow full accessibility, which includes not only a legal authorization to perform certain rights, but also the technical possibility to effectively access and reuse material. Then, based on the example of attribution and authorship requirements for reproduction and performances of works and derivative works, it will be examined what technical infrastructure may better support the enforceability of these licensing terms, namely a framework automating certain actions and pedagogy tools.

From legal accessibility to technical accessibility and technical support of open content licenses, this article illustrates the intricate relationship between law and technology in the realm of copyright and focuses on access and authorship, two fundamental elements of (free) culture and (open) science.

Keywords: Creative Commons licenses, open access, open data, free culture, technical accessibility, attribution, authorship, credit, derivative works.

1. Introduction

Creative Commons (CC) offers a set of copyright licenses for authors who want to grant flexible rights to the public. The licenses are generated through a web interface which delivers a piece of HTML code to the user. The licenses have various formats which are linked to one another. The HTML code describes a license button, which is a logo with an embedded link to the human-readable Commons Deed summarizing the provisions developed in a Legal Code, the actual text of the license. The machine-readable layer of the licenses makes it possible for search engines to locate content marked with CC metadata. All licenses have in common the requirement that the licensee attribute the work and retrain a link to the CC license for any use or redistribution. Various options allow the author to retain or grant commercial rights and the rights to make derivatives. The

licenses combine several elements around core clauses expressing the rights granted and restrictions which may apply, as well as general conditions.

In a first part, I will discuss which legal and technical conditions make it possible to reuse works licensed under a CC license and what could be added to make reuse more effective. In a second part, I will explain that reuse requires correct attribution and how this task could be better supported.

2. The conditions of technical accessibility

Open content licenses intend to facilitate sharing and reuse. All CC licenses authorize the public to reproduce and publicly perform their work, including in a collection (a selection or an arrangement of several unmodified works, such as an encyclopedia or an anthology). The rights granted “may be exercised in all media and formats (and) include the right to make such modifications as are technically necessary to exercise the rights in other media and formats”¹.

Some CC licenses allow adaptations, which “means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, arrangement of music or other alterations of a literary or artistic work, or phonogram or performance and includes cinematographic adaptations or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original”, including the synchronization of music on moving images.

Therefore, it is expected that the original work can technically be modified, not only to make adaptations, but also (and anyway in all the licenses) to transfer it to other medias or formats, or to reformat it to include it in a broader collection. This permission requires delivery in a format that effectively enables reuse. After defining the notion of open access and proposing to include technical accessibility, the concept will be applied to scientific publications, scientific databases and cultural works.

¹ All definitions and excerpts of the CC licenses come from the legal code of the version 3.0 of the licenses, for instance <http://creativecommons.org/licenses/by/3.0/> for the CC Attribution license.

2.1. DEFINING OPEN ACCESS

Culture and science are being built incrementally. Artists take inspiration by others' works, and scholars also reuse previous articles and data. All aim to broadly disseminate new culture and the knowledge they create into society. The Free Culture movement "promotes the freedom to distribute and modify creative works using the Internet as well as other media"². "A free culture is one where all members are free to participate in its transmission and evolution, without artificial limits on who can participate or in what way."³ The Open Access (OA) movement seeks to take advantage of the opportunities offered by digital publishing and distributing to share scientific results more quickly. Both movements aim to facilitate education, culture and access to knowledge, and CC licenses are one of the tools towards these objectives. What steps can further free culture, open education, open science and scholarship? Should licenses simply ensure access without a fee while granting some legal rights, or should the licenses do more to improve that access, such as including technical capabilities for finding, extracting, modifying, editing, remixing, annotating, compiling and otherwise tweaking the content in order to make better use of it? In order to define OA, this subsection draws examples from the situation of OA in respect to scientific publications, which has a longer history than Open Science or Open Data and Free Culture, which will be discussed in the next subsections.

Based on the Budapest Open Access Initiative definition for Open Access, "free availability (...) without financial, legal, or technical barriers"⁴, three categories are fundamental for OA material. They constitute a typology to define the different forms of OA: **economic OA**, **legal OA** and **technical OA**⁵. Usually, the emphasis is put only on the two first categories and I propose to give a specific attention to the technical barriers to OA which are often hidden or neglected.

2.1.1. *Economic OA*

Research available only for a fee cannot be read by researchers from financially disadvantaged institutions and countries where libraries cannot

² http://en.wikipedia.org/wiki/Free_culture_movement

³ <http://freeculture.org/>

⁴ <http://www.sorors.org/openaccess/read.shtml>

⁵ The three following subsections defining legal, economic and technical open access come from Mélanie Dulong de Rosnay (2008), « Opening Access in a Networked Science », in *Publius Project, Essays and conversations about constitutional moments on the Net collected by the Berkman Center*, June 2008. <http://publius.cc/2008/06/13/melanie-dulong-de-rosnay-opening-access-in-a-networked-science/>

afford the subscription to a particular journal or online database. The public also will likely not afford these articles either. Economic OA grants basic access rights by making articles and data available for private reading.

Economic barriers to access can be waived through different options. Publishers can issue OA journals that do not charge their readers and they can develop alternative publication models: this is the golden road to OA. Authors can also self-archive their articles in pre-print or post-print versions in an institutional repository. Many non-OA journals allow authors to do so, and this is Green OA. Several policies are available for those authors who want to but cannot. Authors may add a contractual opt-out clause⁶ to their publishing agreement to retain some of their rights. Finally, universities and research funders may mandate the archiving of articles in OA repositories.

2.1.2. *Legal OA*

Legal OA is an additional condition that allows redistribution, and it goes beyond the removal of financial barriers to accessing and reading. Removing permission barriers grants the public rights to use material beyond simple access. Like economic OA, legal OA, or “Permission-barrier-free” scholarship relies on contractual agreements. Authors must indicate that they are publishing their output without legal restrictions. Otherwise, third parties will not be aware that they may have additional permissions beyond the right of reading. Without an explicit declaration that additional rights are granted to the public, the right to copy, distribute and make derivatives may be impeded by transaction costs associated with permission requests. Libraries, professors, and other curators and aggregators may wonder if they can reproduce, translate, and redistribute material on websites or in course packs without an expensive rights clearance process. Adding a clear license to a journal, repository, or conference website will allow creative and confident usages. The Creative Commons Attribution license⁷ complies with the Budapest Open Access Initiative definition and makes legal OA a reality. However, other Creative Commons licensing options that reserve commercial rights and derivative rights do not comply with this definition and can not lead to legal OA. In these instances for example, one may redistribute legal OA articles only for non-commercial purposes, or one may not translate them or distribute derivative works without additional authorization.

2.1.3. *Technical OA*

Just as in the case of price and rights clearance, technology can create barriers to access, redistribution and reuse of articles and data. But

⁶ For instance <http://scholars.sciencecommons.org/>

⁷ <http://creativecommons.org/licenses/by/3.0/>

technical choices can also help remove barriers. Technical OA should ensure that materials can be actually and effectively reused, mined, processed, aggregated, integrated, and searched by both humans and machines. Technical barriers can include the following: protection measures that prevent copying, compulsory registration before download, design features that add hidden costs to search and processing, complexity of all sorts prior to full accessibility of the content in a data format allowing any sort of processing. For example, it can be more or less easy to interact with a document because of the publication format. HTML pages are more convenient to browse a large amount of articles compared to PDF files which require download. HTML and wiki allow comments and editing; two-column articles are difficult to read quickly on most screens but are the norm for scientific articles. Poor indexing or lack of metadata also prevent some modes of use.

The opening of this triple architecture of market, law and technology allows broader and better access. More and more journals and book editors are becoming aware of OA's social benefit and potential impact on innovation and aim to share their results. If they wish to do so, they should make sure that not only economic and legal, but also technical restrictions have been effectively removed, so that researchers and the public can not only access, but also redistribute and reuse materials in any way, including ways that initial creators had not considered.

2.2. ACCESSIBILITY FOR SCIENTIFIC DATABASES

Removing technical restrictions to full OA has a different meaning for scientific publications than for scientific data, and data curators may wonder what accessibility or open formats mean for scientific databases.

Contractual requirements such as Creative Commons Attribution policy⁸ and the complexity of these requirements⁹ constitute a legal barrier to

⁸ Science Commons suggests to distribute data under simple and understandable terms as close as possible to the public domain, free of copyright, contractual, database and other controls. Nguyen Tinh, «Freedom to Research: Keeping Scientific Data Open, Accessible, and Interoperable », Science Commons Reading Room. <http://sciencecommons.org/wp-content/uploads/freedom-to-research.pdf>

⁹ Legal simplicity and predictability can be achieved by waiving copyright and other contractual restrictions, allowing data integrators to reuse, modify and redistribute large datasets, towards the freedom to integrate according to Wilbanks John, "Public domain, copyright licenses and the freedom to integrate science", *Journal of Science Communication*, volume 07, issue 02, June 2008.

downstream reuse of data. But legal accessibility is not the only hurdle to data integration. Technical accessibility as defined above should be ensured in order to allow scientists to download data easily and use them in any way, including ways that the initial creators, curators and contributors had not considered.

The objective of the research¹⁰ presented hereafter has been to assess the accessibility of databases by analyzing their access interfaces and their reuse policies. Databases' openness will be measured by analyzing a set of technical access interfaces and legal terms of use. A taxonomy of technical and legal restrictions applicable to databases in life science will be presented and used to assess a sample of databases. Based on these criteria, I propose a set of questions for database curators to assess their own data's technical and legal openness. It intends to help to define what can be changed or specified in open content licenses to better support full accessibility in the context of databases of scientific data.

2.2.1. *The design of a taxonomy*

This research started with analyzing the terms of use for databases from the Molecular Biology Database hosted by the Nucleic Acids Research Journal (extend link to whole name), and assessing them regarding open access criteria as described in the Science Commons Open Data Protocol¹¹. A sample of policies has been retrieved and analyzed. The next step identified barriers to open access and reuse of data based on these database policies and built a taxonomy of restrictions. These restrictions can be of legal or contractual nature, but they can also be technical, e.g. the impossibility of downloading the whole database if its results can be accessed only through a field-based search. A systematic analysis of more database policies hosted by the Life Science Resource Name ([LSRN](#)) Schema registry allowed us to confirm this taxonomy and to refine it by adding other terms.

¹⁰ This section reuses substantial parts of a paper the author wrote with Shirley Fung entitled « Legal and Technical Accessibility for Life Science Databases », Proceedings of the Second Communia Conference: Global Science & Economics of Knowledge-Sharing Institutions, Torino (Italy), 28-29-30/06/2009 available at: <http://www.communia-project.eu/node/333> which was itself developing a preprint (Mélanie Dulong de Rosnay, “Check Your Data Freedom: Defining a Taxonomy for Access and Reuse of Life Science Data”, *Nature Precedings*, July 2008) the author wrote as the output of a fellowship at Science Commons on a research project developed under the auspices of the Science Commons Data project and building upon the Science Commons Open Access Data Protocol proposing requirements for interoperability of scientific data available at:

<http://sciencecommons.org/projects/publishing/open-access-data-protocol/>

¹¹ <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>

Technical and legal accessibility conditions as well as restriction values will be defined. The purpose of identifying controls and restrictions applicable to databases is twofold. First, it will enable the understanding of the terms of use and other requirements governing the access to molecular biology databases especially identifying the control that prevents the free sharing of data. Second, these restrictions will be clustered into classes, making it possible to systematize the analysis of databases and to easily identify the data that can be reused by the scientific community.

Two types of control can be exercised on databases: technical restrictions embedded in the design of the database, and legal restrictions expressed in the terms of use.

Technical restrictions affect databases that cannot be searched or processed in any possible way. Technical openness is ensured by the possibility of downloading the whole dataset and reusing and integrating data, in the same way the Science Commons Neurocommons project provides a data mining platform allowing machine-readable representation and interpretation of data, or that Basic Local Alignment Search Tool (BLAST) finds similarities between sequences. Semantic web processing applied to scientific data should improve the way science is performed and allow network effects by connecting knowledge from various datasets. Databases that require registration before access, or offer only a batch processing or a query-based mechanism to retrieve data after a specific search, do not comply with the technical requirements necessary to make data open.

Terms of use, licenses and access policies are legal texts describing authorized and unauthorized usages. The legal rules are expressed by the entity distributing a product such as software or scientific data. The infringement of these self-declared rights can lead to lawsuits. Terms of use can be difficult to understand, even for lawyers, while scientists need to know quickly whether they can use a dataset.

Therefore, a set of questions has been designed to understand whether databases are, in fact, fully accessible and whether the data can be reused, redistributed and integrated.

Technical accessibility

Downloadability

Is there a link to download the whole database?

YES or NO

If YES, include the URL

Batch

Is it possible to access the data through a batch feature?

YES or NO

Query

Is it possible to access the data through a query-based system?

YES or NO

Registration

Finally, is registration compulsory before downloading or accessing data in the ways described above?

YES or NO

Legal accessibility*Terms of use*

Does the database have a policy?

YES or NO

If YES, include the URL and assess whether the policy authorize reuse, redistribution, integration

Are there any restrictions on the right to reformatting and redistributing?

If NO

If YES, which restrictions?

Fields to describe restrictions are

Attribution Contractual Requirements

Non-Derivative Use

Non-Commercial Use

Share Alike

Others (to be described).

Figure 1. Set of questions to process databases

The questions in Figure 1 allow the processing of databases. A subsequent database has been developed with information describing databases technical and legal accessibility.

Five answers can be provided to these questions and together constitute a taxonomy to assess technical and legal openness, as presented in Figure 2 below.

1. DOWNLOADABILITY

The website provides a file transfer protocol or a link to download the whole dataset without registration.

The ability to download the whole dataset without registration constitutes the double requirement to be considered as technically accessible.

<p>2. TECHNICAL RESTRICTION: the database can be accessed only through registration, batch or query-based system. Technical accessibility is not achieved.</p> <p>3. PUBLIC DOMAIN POLICY: the website provides simple and clear terms of use informing users that the data are in the public domain. Data are thus free to integrate. Legal accessibility is achieved.</p> <p>4. NO POLICY: the website does not provide terms of use. Legal accessibility is not achieved.</p> <p>5. LEGAL RESTRICTIONS: the terms of use impose contractual restrictions, such as heavy contractual requirements for attribution, limitation to non-commercial usages, prohibition to modify data, or other constraints on their redistribution or modification. Legal accessibility is not achieved. The data are not free to integrate.</p>
--

Figure 2. Databases qualification

2.2.2. Databases analysis according to the taxonomy

Samples of the Nucleic Acids Research (NAR) Molecular Biology Database Collection MBDC¹² and Life Science Resource Name (LSRN)¹³ Schemas databases have been analyzed to define the taxonomy. Then one third of the LSRN databases (60 databases) have been systematically analyzed. A subsequent database¹⁴ has been created, gathering for each of these databases:

- The name and URL of the database,
- URL of the download page and URL of the terms of use,
- Extracts of the terms of use for further review and comments,
- Values for technical accessibility and legal accessibility features as described in Figure 1.

Technical openness

Four values have been identified to assess technical accessibility: Downloadability, Batch features, Query-based system and Registration.

¹² <http://nar.oxfordjournals.org/cgi/content/full/gkm1037/DC1/1>

¹³ <http://lsrn.org/lsrn/registry.html>

¹⁴ A user interface has been built by Shirley Fung using PHP and MySQL to host the dataset assessing databases technical and legal accessibility. It is available at <http://labs.creativecommons.org/demos/mbdb/>

The only combination qualifying the database as technically open is the ability to download without registration. Indeed, registration before access and the possibility to perform only batch or query-based searches prevents automated data mining. However, it can be useful to have access to several systems to retrieve and analyze data. Therefore, the database indicates whether it is possible to retrieve data also through batch and query in addition to download.

Another technical restriction that has not been analyzed is the presence of standardized annotations or comments allowing users to understand data collected by others. This feature has been disregarded because of a lack of expertise to assess the relevance and quality of annotation for external reuse.

Legal openness

Values that have been used to define legal accessibility are the following: **Policy Available, Public Domain Policy, Attribution Contractual Requirements, Non-Derivative Use, Non-Commercial Use, Share Alike and Other (to be described).**

In order to be open, a database must have a policy, and this policy must not impose any restriction to the redistribution and the modification of data. The absence of any terms of use or policy on the database website could imply to some people that, in the absence of any expressed restrictions, data are free. But rights unknown to the user might be applicable by default. Indeed, the Science Commons Protocol states that “any implementation MUST affirmatively declare that contractual constraints do not apply to the database.” Policies should be clear and have only one possible legal interpretation. The absence of a clear and understandable policy is equivalent to the absence of a policy because it leads to legal uncertainty.

Legal restrictions to redistribute and modify of data can be diverse. Four values have been identified, corresponding to Creative Commons licenses options: Attribution Contractual Requirements, Non-Derivative Use, Non-Commercial Use, Share Alike. However, the definition for these legal restrictions in the context of this research is broader than the Creative Commons definitions.

The Attribution requirement may constitute a restriction on the reuse of data. Instead of strong contractually binding requirements on how data should be attributed, a request of acknowledgment according to scientific norms should be sufficient. According to the Protocol, “any implementation SHOULD define a non-legally binding set of citation norms in clear, lay-readable language”. Furthermore, “community standards for sharing publication-related data and materials should flow

from the general principle that the publication of scientific information is intended to move science forward. An author's obligation is not only to release data and materials to enable others to verify or replicate published findings (as journals already implicitly or explicitly require) but also to provide them in a form on which other scientists can build with further research."¹⁵

The Non-Commercial and Non-Derivative requirements prevent many types of data use. They are defined as restrictions based on the commercial nature of the user or of the usage, and as restrictions on the distribution of modified versions of the database.

The Share Alike requirement is present in the original taxonomy. This option requests modifications to be offered under the same open terms and should encompass all copyleft policies. No policy in the analyzed sample contains this requirement.

Other possible restrictions may affect terms of use. For instance, an embargo on publishing before the data producer, the existence of patents and the absence of warranties against third-party rights are legal restrictions which have not been taken into account in this first analysis.

In many cases, the database is offered with no restrictions in place by the database curator, but nevertheless without warranties on the legal status of the data submitted by contributors. Data may contain elements protected by copyright or any applicable right. The database curator did not clear the rights, or did not request from the contributors a rights waiver or no rights assertion before data upload, or does not want to be held liable in case the previously described processes would present a failure. This warranties disclaimer can be seen as a hurdle to the usage of these data. Both uncertainty for the end-user and absence of responsibility for the curator might be avoided by offering contributors a seamlessly integrated data sharing agreement prior to submission. Although this procedure might disincentive some contributors, the burden of checking the legal status of data and avoiding possible claims by third parties should not rely on the data user, forcing her to hire a lawyer. Besides, these disclaimers do not identify which data are free and which parts of the database might be copyrighted or covered by other rights.

2.2.3. Results

Databases which can be considered legally and technically open, and compliant with the Science Commons Open Data Protocol, are those that are downloadable without prior registration and under a simple policy close

¹⁵ Board on Life Sciences (BLS), *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences* (2003).
<http://books.nap.edu/books/0309088593/html/R1.html>

to the public domain. The impossibility or the difficulty of downloading and reformatting the dataset does not fulfill technical accessibility requirements. Databases available only through batch or query interfaces are not considered technically open, but those offering these features in addition to downloadability will be compliant.

Besides databases created by the National Center for Biotechnology Information (NCBI) and the European Bioinformatics Institute (EBI), only a couple of databases among the 60 first schemas of the LCRN registry analyzed sample can be considered as both technically and legally open, without restrictions.

2.2.4. Checklist to assess databases openness

The following checklist may assist data curators in opening their data, and to make sure that the database's design and terms of use will allow others to access, reuse and build upon their data. All answers should be positive.

A. Check your database technical accessibility

- A.1. Do you provide a link to download the whole database?
- A.2. Is the dataset available in at least one standard format?
- A.3. Do you provide comments and annotations fields allowing users to understand the data?

B. Check your database legal accessibility

- B.1. Do you provide a policy expressing terms of use of your database?
- B.2. Is the policy clearly indicated on your website?
- B.3. Are the terms short and easy to understand by non-lawyers?
- B.4. Does the policy authorize redistribution, reuse and modification without restrictions or contractual requirements on the user or the usage?
- B.5. Is the attribution requirement at most as strong as the acknowledgment norms of your scientific community?

Figure 3. Database openness checklist

2.3. TECHNICAL ACCESSIBILITY FOR LITERARY AND ARTISTIC WORKS

The absence of economic and legal restrictions expressed in clear and simple terms of use is not enough to ensure full accessibility to scientific articles or data. Distribution should ensure that materials can be effectively reused and processed by humans and machines. Several features typical of a bad design should be avoided in order to facilitate data mining and further aggregation and integration in collections and derivative works:

registration, abuse of PDF or format which is cumbersome to process and edit, difficulty in downloading an entire set or the content of a website within a few clicks.

Can these requirements be useful beyond science? Are these remarks applicable to literary and artistic works? Could the CC licenses encourage technical accessibility? After having defined what is technical accessibility for science, I will try to define how the technical barriers and requirements above can be transposed to other works.

Technical accessibility and the ability to manipulate software are conditioned by the release of the source code in an open format and OA to the relevant documentation. The GNU Free Documentation License¹⁶ is the standard license for software documentation. It contains in its first clause a definition of open format crafted for textual software documentation:

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

This definition allowing human editing and machine processing applies to textual media in the current state-of-the-art text editors and technical standards. OGG is the free and open format for audio while MP3 is a de facto distribution standard. Music players and editors necessary to reproduce, perform and adapt MP3 files may require the payment of a fee at some stage, which may conflict with economic OA at some point.

But the situation is more complicated for media types other than text. Again taking the example of music, having an audio file in a free and open

¹⁶ <http://www.gnu.org/copyleft/fdl.html>

format may not be sufficient to remix it. Instructions such as information enclosed in a MIDI file and other data such as music notation or explanation for performers could, when they exist, also be released. Some of these issues can be solved through project-oriented online communities which would encourage or require uploading complete project-files in addition to the media.¹⁷

Open media means more than distributing a file in non-proprietary format under an open content license. CC licenses, as we will see in the coming section, contain very detailed requirements on the way to attribute authors. They could also contain requirements to facilitate technical accessibility and allow a true remix culture. Guidelines for each community or media type could be developed and, without being as specific as the GFDL clause, a provision could request the licensor to release the work in a format suitable for manipulation and with the information necessary for its manipulation in a reasonable manner appropriate to the media. This implies a different perspective than the current CC approach which places restrictions on the licensee rather than on the licensor.

3. Technical contributions to authorship

Now that I defined what technical accessibility could mean for free culture and open media to ensure full access and facilitate reuse beyond a CC license grant, I propose to accompany open content licensing with a technical framework facilitating authorship and attribution. The concept of attribution is central to copyright from a civil law country perspective with strong moral rights, but not exclusively. Citing the author is a social norm beyond legal and contractual obligations. The Attribution element is standard in all the CC licenses; they all require the original author to be credited for her work when copying, performing or remixing it.

3.1 THE CC ATTRIBUTION CLAUSE

The Creative Commons Attribution provision addresses not only the name of the author, but also the name of one or several individuals or entities who can be not only authors or performers but also licensors, rightholders, publishers, sponsors, etc. as well as the URI associated to the work. The attribution provision is expressed in the CC Commons Deed, the human-readable summary, as follows:

¹⁷ Cheliotis Giorgos, "From open source to open content: Organization, licensing and decision processes in open cultural production", *Decision Support Systems*, Volume 47, Issue 3, June 2009, p. 229-244.

“Attribution — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).”

Besides, all the licenses require the user to also include the license when they reuse the work:

“Notice — For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.”

The right to make derivatives granted in some of the licenses is stated as:

“to Remix — to adapt the work”

The attribution provision has a long definition with specific requirements located in three subclauses:

1. In the license grant clause for the licenses authorizing adaptations to condition the exercise of this right to the identification of the changes made to the original work¹⁸,
2. In the second subclause of the restrictions clause as a positive obligation of the licensee to attribute the author or licensor as she requests, including the attribution of adaptations if they are authorized, and the way to exercise this obligation,
3. And at the end of the first subclause of the restrictions (4.a.) as a negative obligation to remove upon request of the licensor such attribution from collections and adaptations to the extent they are authorized.

The text, which varies among licenses authorizing adaptations and licenses that do not, reads as follows, with provisions related to derivatives in italic and a modified layout and order of the excerpts to present them in the order they are to be exercised, starting with requested attribution, including for adaptations, followed by non endorsement and unwanted attribution requirements:

<p>“If You Distribute, or Publicly Perform the Work or any Adaptations or Collections, You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied,</p>

¹⁸ This provision “The original work was translated from English to Spanish” could be clustered with the next one “French translation of the Work by Original Author”: even if the first is addressing the original work and the second the author of the original work.

and/or if the Original Author and/or Licensor designate another party or parties (e.g., a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties;

(ii) the title of the Work if supplied;

(iii) to the extent reasonably practicable, the URI, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work; and

(iv) *consistent with Section 3(b), in the case of an Adaptation, a credit identifying the use of the Work in the Adaptation (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author").*

(in clause 3 License grant)

to create and Reproduce Adaptations provided that any such Adaptation, including any translation in any medium, takes reasonable steps to clearly label, demarcate or otherwise identify that changes were made to the original Work. For example, a translation could be marked "The original work was translated from English to Spanish," or a modification could indicate "The original work has been modified.";

The credit required by this Section 4(c) or 4(d) may be implemented in any reasonable manner; provided, however, that in the case of a *Adaptation or Collection*, at a minimum such credit will appear, if a credit for all contributing authors of the *Adaptation or Collection* appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors.

For the avoidance of doubt, You may only use the credit required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.

If You create a *Collection*, upon notice from any Licensor You must, to the extent practicable, remove from the *Collection* any credit as required by Section 4(b), as requested.

If You create an Adaptation, upon notice from any Licensor You must, to the extent practicable, remove from the Adaptation any credit as required by Section 4(b), as requested."

Figure 6. CC licenses Attribution provisions

To summarize, the license allows the licensor to require from the licensee a particular way to attribute the work by citing:

- The name of the author, licensor or any party,
- The title of the work,

- The source URL (not URI?) of the work,¹⁹
- For derivatives, a credit identifying the original author, the use of the original work and changes which have been made.

3.2 FACILITATING ATTRIBUTION AND AUTHORSHIP

The Attribution requirements are difficult to fulfill. An initial solution could be to simplify the wording of the Attribution clause in the CC licenses. Meanwhile, this final section describes difficulties raised by the high standard of attribution in the CC licenses and proposes possible solutions for better compliance.

The licensor may require that these elements be cited to the extent she supplies them (except for the last one requiring to identify changes made to the original work in a derivative because it is not possible). It is not clear how the licensee should fulfill this obligation in case no or insufficient information has been provided by the licensor who may not have the skills or the energy necessary to express this information. Sometimes the original licensor did not correctly, fully or entirely express attribution of the original work in the first place.

Some websites provide useful guidelines²⁰. The standard of attribution is “a reasonable manner” except for adaptations and collections, where it should follow as a minimum the attribution standard of the other components.

¹⁹ But not the source URL of the original work for derivatives, which could be useful, as allowed by the fields on the license chooser interface: <http://creativecommons.org/choose/>

²⁰ See for example the attribution policies or guidelines published by Global Voices hosting articles authored by bloggers and translated by others, providing a recommended model to attribute as well as expressing the wish to have the logo of the institution included in addition to the name of the author and the hosting website, corresponding to the “publishing entity” as defined by the CC license attribution clause: “Please note: in the case of images and multimedia we have sourced from others you need permission to republish from the creators, as they may have different copyright terms from Global Voices. Please include a link to the original article. An example:

This article by Jane Doe was originally published by Global Voices Online, a website that translates and reports on blogs from around the world.

If you want to make us extra happy, please include our logo in the attribution. To make it easier, you can copy and paste the code below to make the image appear in your blog or website.”

<http://globalvoicesonline.org/about/global-voices-attribution-policy/>

In the case of complex mash-ups of artistic and literary works, it is sometimes difficult to know whether a final collage will be considered as a collective work. This qualification matters because the CC licenses attribution provision stipulates that “in the case of an Adaptation, a credit identifying the use of the Work in the Adaptation (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author"” is needed. Attribution must be expressed in a specific way and conform to the format, medium or means employed to convey the work. Finally, the copyright notice must solely express attribution: no connection, sponsorship or endorsement by the original author may be asserted by the licensee reusing the work. The licensee should not use the credit to imply the author, licensor or party is endorsing the licensee or her use of the work. The licensee must be ready to remove the credit from adaptations and collections upon request from the licensor. This requirement raises practical questions. The licensor may never notice the work, or notice it late and make it impossible for the licensee to remove credits on works which have already been circulated, shared and reused.

Attribution requirements are not only considerable for complex works reusing numerous prior contributions or collections and successive adaptations which might be difficult to trace. Attribution is often not fulfilled by licensees because they do not know how to proceed. They may provide the name of the author, but not the link to his webpage nor to the CC license, or mention that the work is under a CC license and link to its source without crediting the author. A common example where attribution is badly handled is found in newspapers reusing Flickr photos without proper credit, possibly creating a copyright infringement, a violation of the license (and thus the termination of the grant), and the impossibility of subsequent users properly attributing the work if they wish to reuse it or incorporate it in their own work.

An initial technical solution could be to better assist the licensor and the licensee with filling in adequate fields with appropriate information. This task can be facilitated by applications that would automate the process for both 1) licensors, who when selecting a license to apply to their file²¹, should enter correct and complete data in the license chooser interface which already contains fields for optional additional information, and 2) licensees when editing and redistributing a modified work. It is already possible for the licensor to indicate the following in the CC metadata: 1) the format and the title of the work, 2) the name users of the work should give attribution to, 3) the URL users of the work should link to, the source work URL and 4) an URL for additional permission. If this option was

²¹ <http://creativecommons.org/choose/>

more widely used and further developed, licensees would receive proper information in the work's metadata, and further applications or editors plugins could help provide correct attribution when they redistribute the work or reuse it otherwise. A simple specification of attribution elements²², would help authors and licensors to be attributed the way they are entitled to request, and to help licensees to respect these requirements. Then, attribution elements would follow the work along its life-cycle.

The burden of legalese requirements should be kept minimal in order to not deter creativity and discourage innovation: the goal of open content licenses is to make circulation, use and reuse easier, not to add complexity for licensors to indicate how they want to be attributed and for licensees to respect these contractual expectations.

A second solution also involving the support by technology is the development of more tutorials, comics or games describing the role of authors and reusers and conveying authorship ideas which are behind the attribution clause. Explaining how to attribute a work can also be seen as the first step to teach how to create, use and reuse creative works, and give a sense of what constitute an act of authorship when one creates or modifies someone else's work. Beyond sophisticated legal clauses and technical applications to convey attribution elements, users may lack necessary media literacy and copyright law skills to understand what is authorship and what constitutes an adaptation. The New Media Literacy team²³ developed an attribution module within an online learning environment the Learning Library²⁴. Challenges, or sequences of a game to teach media literacy and participation, involve activities around copyright, fair use and the CC licenses.

With the NML team, solving the difficulties of understanding authorship and reuse were identified as prerequisite to using a CC license. To indicate proper attribution for both potential licensors and licensees, I suggested this simple three-step pattern of reuse:

1. An author creates a work.
2. Someone else modifies this original work.
3. The result constitutes a new work.

²² Which can take the format of trackbacks and of RDF tags supported by the CC Rights Expression Language (ccREL) described at <http://wiki.creativecommons.org/CcREL>

²³ <http://newmedialiteracies.org/>

²⁴ <http://newmedialiteracies.org/library/>

Understanding authorship should help to express attribution as follows: “This is a work by [name of the author of the adaptation] that is [modification action] from [title and link to original work], by [name of the original author].”

Conclusion

Works and data made available under a CC license may require more freedoms than the licenses offer to maximize possibilities of reuse and remix. They should be made available in ways and formats that technically enable modifications, editing and processing and the licenses could also foresee to waive technical restrictions. Technology can also help ensure attribution accompanies media files. Pedagogy is also necessary for both licensors and licensees to understand authorship and build a free culture.

Acknowledgements

The author is grateful to Shirley Fung for her work on the databases user interface, to Shunling Chen, Erin Reilly, Anna van Someren, Victoria Stodden, John Wilbanks and Carolina Rossini for many valuable discussions, and to Michelle Thorne for her useful comments.

References

- Cheliotis G. (2009), *From open source to open content: Organization, licensing and decision processes in open cultural production*, Decision Support Systems, Volume 47, Issue 3, June, pp. 229-244.
- Dulong de Rosnay, M. (2008a), *Opening Access in a Networked Science*, in Publius Project, Essays and conversations about constitutional moments on the Net collected by the Berkman Center, June.
- Dulong de Rosnay, M. (2008b), *Check Your Data Freedom: Defining a Taxonomy for Access and Reuse of Life Science Data*, Nature Precedings, July.
- Dulong de Rosnay M., Fung S. (2009), *Legal and Technical Accessibility for Life Science Databases*, Proceedings of the Second Communia Conference: Global Science & Economics of Knowledge-Sharing Institutions, Torino (Italy), June, 28th-30th.
- Nguyen, T. (2008), *Freedom to Research: Keeping Scientific Data Open, Accessible, and Interoperable*, Science Commons Reading Room. <http://sciencecommons.org/wp-content/uploads/freedom-to-research.pdf>
- Wilbanks J. (2008), Public domain, copyright licenses and the freedom to integrate science, Journal of Science Communication, Vol. 07, Issue 02, June 2008.

URLs were last accessed by 26 November 2009.