Copyright challenges in the transition to FAIR research data at UiO - Note from a QualiFAIR working group

Agata Bochynska, Rebecca Josefine Five Bergstrøm, Alexander Refsum Jensenius

15.02.2024

Recommendations

The working group sees the need for:

- developing information pages about copyright and licenses for researchers and students
- recommending some licenses for different data types (some open and some restrictive)
- developing systems that help with license management (on everything from individual files to databases)
- developing copyright courses for employees and students at all levels
- raising copyright awareness among managers at all levels
- clarifying the distribution of responsibilities and who acts as the first, second and third line of contact for both students and staff

---

Background

The competence hub QualiFAIR (2021-2024) at the University of Oslo (UiO) has been focusing on how qualitative, context-sensitive and personally identifiable data can be made FAIR (Findable, Accessible, Interoperable, Reusable). The Hub has been organized as a network of academic and technical-administrative staff from various disciplines, including anthropology, political science, medicine, linguistics, psychology, musicology, theology, and education. The hub's main goal has been to collect and establish good practices, including routines, procedures, and concrete guidance, for supporting the sharing and reuse of qualitative data, aimed at researchers, master's/PhD candidates, and research support staff at all departments and faculties at UiO. This note summarizes the findings of the QualiFAIR's Copyright Working Group.

Process

The working group had representatives from various disciplines and units and also included additional resource persons from within the organization. The following groups and disciplines were involved in the work in various ways: musicology (Alexander Refsum Jensenius, IMV/HF), linguistics (Kristin Hagen, ILN/HF), media studies (Arnt Maasø, IMK/HF), school research (Torgeir Christiansen, ILS/UV), the University Library (Agata Bochynska, Matthew Good, Solveig Sørbø and Rebecca Josefine Five Bergstrøm, UB), the central administration (Margaret Louise Dyrnes Fotland and Jostein Greve, FIADM), and IT security (Vilde Sørbø Nenseth and Isak Falch Alsos, IT Department).

Three thematic meetings were held in the academic year 2022/23:

- 21.10.2022: Discipline-specific issues: RITMO (IMV/HF), TekstLab (ILN/HF), IMK (HF), ILS (UV).

- 20.01.2023: The role of the research support units: USIT, UB, FIADM.
- 20.03.2023: Summary and discussion of the follow-up plans.

The group has also mapped the status at UiO through the QualiFAIR network and a survey. Preliminary findings from the work were presented at a well-attended breakfast seminar on April 18th, 2023.

Basics of copyright

Copyright is the rights a creator of an intellectual work has to the work. This includes the exclusive right to dispose of the work, including production and making it available, as well as adaptations and translations. Copyright is regulated in Norway by the Copyright Act. Copyright is limited in time, as a general rule, to 70 years after the end of the year of the author's death.

The Copyright Act protects representations, literary or artistic works. Thus, the concept of the *work* is delineated against facts, principles, methods, mathematical concepts, etc. The prerequisite for copyright protection is a practical, creative effort and individual intellectual activity. Copyright protection covers not only the work in its original form but also in "altered form, in translation or adaptation". Many copyright disputes are related to reproduction or copyright infringement when producing new intellectual works. In such cases, the question is whether the new intellectual work has a "threshold of originality", i.e., literary or artistic independence. In this context, we mainly mean cases of reuse of material where it is clear that someone else has the copyright.

Through their work, researchers also get their own copyright to the material they produce, such as articles, books, and presentations. The rights become more complex when handling datasets and databases, especially if they are based on the work of others.

In our work, we have seen that it is useful to divide the problem into two:

- Other's copyright: What can I do with other's material?

- Own copyright: What can others do with my material?

In the following, we will give a brief summary of the problems.

General observations

The review has shown that, in general, there is little expertise on copyright among academic and technical-administrative staff at UiO. They also don't know where to turn for help. The problem is exacerbated by the lack of central guidelines and recommendations. This is in stark contrast to the growing knowledge and awareness of privacy, which is due to a large focus on privacy at many levels in the organization recently, but correspondingly little on copyright.

Although there is generally little knowledge of copyright within the organization, specialist expertise exists in several places. There is, of course, a lot of professional expertise on copyright issues at the Faculty of Law, but it is not available to researchers and students at other faculties. In the support apparatus, there is relevant legal expertise at USIT, FIADM and UB. The legal advisors with relevant expertise work largely with privacy issues rather than copyright. The survey and interviews in the QualiFAIR network have shown a need for better coordination between these units, clarification of who should act as the first line of contact for researchers and students, and the second line for more advanced questions.

Other's copyright

The working group has mainly focused on qualitative research material. This type of data is stored in many different formats: text, image, sound, video, and tables. The collected examples cover the subject areas of music research, linguistics, media studies, political science, sociology, and school research. The issues are different, but all related to copyright connected to sharing and reusing research data. Here are some examples from various fields:

- **Music research**: The researchers make recordings of concerts where different pieces are played, all of which can have several different authors: composers, orchestrators (who notate other people's music), and lyricists. During some concerts, the musicians improvise and thus become composers themselves. Several concerts also have elements of dance and visual art that have their own creators. In addition, rights can be so-called related rights, and management organizations can manage both the copyrights and the related rights. Further complicating matter is that separate rules apply to video recording of a concert ("synchronization" of sound and image). A separate summary (in Norwegian) has been prepared describing these problems in more detail.

- **Linguistics**: Many linguists need to share data they collect and use in research, for example, small and large corpora based on fiction and other copyrighted texts. It is possible, but very time-consuming, to make agreements with each individual rights holder (author, translator, publisher, etc.). An alternative solution is to apply to the Ministry of Culture in Norway for the production of copies for research purposes in accordance with the regulations of the Intellectual Property Act (FOR-2021-08-26-2608, in Norwegian) § 4. These copies will only be used internally in the given research institution, and sharing them in the future will not be possible.

- **Media studies**: Researchers and students collect and use various media texts for research. These can contain different media types: still images, screenshots, sound, music, or video. This leads to challenges with potentially many rights holders from different countries and different legislations. Few (if any) have the expertise or resources to trace the origin of these and make agreements with them. Awareness and competence about legal limits for the use of this type of material varies among employees and is particularly low among students who are used to sharing media freely on social media. Many follow these habits when sharing complex texts, music, images, and videos in exam papers and assignments.

- **School research**: An important way of collecting data in school research is video recording in classrooms. The researchers have a strong focus on privacy when collecting data from the classroom (and outside the classroom), but little focus on copyright. The problem is that the video footage captures everything in the classroom, including copyrighted material played in the plenary by the teacher. In addition, the material is displayed on students' personal computers/tablets (audio, video, and images). Even if this material is not the researchers' main interest, it becomes part of the data material they handle.

Most of these examples are based on different media types with different copyright regulations and management. The rights and management can also be spread over different participants:

- Individuals or family: In some cases, the owners are still alive and manage their own rights. When the owner dies, in some cases, the family takes over the rights.

- Publisher or company: In other cases, a publisher or company owns the rights.

- Organizations: Some works have collective agreements managed by organizations, such as TONO for compositions, GRAMO for disc recordings, and BONO for visual works.

Sometimes, it is easy to find out who manages the rights, but often, it can be complex to find the right point of contact to answer the questions about the rights. Many researchers don't even know where to start looking. There is often no simple solution to dealing with complex works, where each right must be clarified separately. This would be a lot of work that most researchers do not have the expertise or time to carry out. In practice, this acts as a barrier to sharing data. Solutions may include a combination of training and better support services as well as policy work to reach collective agreements.

Own rights

UiO has an [intellectual property rights (IPR) policy](#) from 2011 which primarily describes the relationship between the employee's and the institution's rights to research results, also in combination with external collaboration partners. One challenge is that many employees are unaware of the guidelines, even though they are written into the employment contract. Many also do not understand how the policy documents work in practice. The current working group observed that more concrete recommendations and guidance are necessary.

The new IPR policy, which will be prepared at the University of Oslo in the autumn of 2023, will include more in-depth guidelines, mainly related to research results used for innovation. This is outside the focus area of QualiFAIR, which aims to look at the sharing and reuse of qualitative and context-sensitive data. From this perspective, the most important tool is raising awareness about using licenses.

Licenses

Licenses are often put forward as the solution to copyright issues. If everyone had assigned (clear and understandable) licenses to their data, a lot would have been resolved. The challenge is that many licenses can be used when sharing research data. Creative Commons (CC) licenses are widespread, but they were originally developed for creative professions and media and, thus, do not necessarily work well for research data. There are also many types of CC licenses, placing different restrictions on reuse (derivatives, commercial use, etc.). There is a need for recommendations on which licenses should be used for which data types. There is also a lack of general knowledge among researchers and others at UiO about the scope and content of the various licenses.

There is also a need for help assigning licenses to datasets based on others' work that does not have a (compatible) license. In these cases, one might have to consider many different licenses and their conditions for reuse. Here, it may be possible to set up systems that help the user find the right license based on the conditions of the underlying licenses. It is also important to create systems that encourage researchers to license their data from the outset to avoid any future problems with data reuse.