

## **Harmonizing Open Licenses among Online Databases of Enthusiast Communities: Challenges Encountered in the Legal Integration of Databases in the Japanese Visual Media Graph Project**

Magnus Pfeffer, Zoltan Kacsuk, Simone Schroff and Martin Roth

*Putting Open Social Scholarship Into Practice: An Implementing New Knowledge Environments (INKE) Partnership & Canadian-Australian Partnership for Open Scholarship (CAPOS) Online Event*

December 8-9 2021 North America time / December 9-10 2021 Australasia time

[1]

Hello, and welcome to our lightning talk “Harmonizing Open Licenses among Online Databases of Enthusiast Communities: Challenges Encountered in the Legal Integration of Databases in the Japanese Visual Media Graph Project” on behalf of myself, Zoltan Kacsuk, and my co-authors Magnus Pfeffer, Simone Schroff, and Martin Roth.

[2]

We will first introduce some necessary background information on our project, database licensing and open licenses and then discuss our experiences with creating an open and license-harmonized integrated database from datasets compiled by online enthusiast communities.

[3]

While more and more scientific projects aim to create large datasets...

[4]

...there are a number of recurring problems even with these projects created or compiled by scientific communities, with unclear licensing and license incompatibilities being the most obvious.

[5]

How can we then hope to create a unified open database from disparate community databases that were originally never meant to be integrated or to further open science? The Japanese Visual Media Graph is an attempt at answering this question.

[6]

The aim of the JVMG project is to work in collaboration with online enthusiast communities to combine their datasets and to open up this data to scientific research.

[7]

Although we do not have time to discuss the complexity of database licensing issues in detail...

[8]

...here is a quick overview of just how many elements of databases can fall under different legal protection in various jurisdictions.

[9]

Which is still further complicated by the fact that all of these elements can also be protected by contractual agreements, as in the case of the Open Database License.

[10]

Which leads us on to our next topic of open licenses for databases.

[11]

Creative Commons licenses and the Open Database License are the two most common open licenses employed for databases.

[12]

Unfortunately, with the exception of the CC0 (which means public domain dedication) and the CC-BY (requiring only attribution) licenses, the CC licenses cannot be combined with the ODbL.

[13]

Furthermore, even CC licenses are limited in their composability...

[14]

...as indicated in this cross-table.

[15]

Turning to the actual licensing practices of online enthusiast communities...

[16]

...we found a range of approaches, from a complete disregard for licensing issues, to various CC licences, sometimes enforcing the openness of derivative databases as well (usually referred to as copyleft), as well as various unique licenses.

[17]

The overwhelming complexity of copyright for non-experts of the subject is well illustrated by two things. First, by our own struggles with the topic in the JVMG project team (indeed, our license harmonization work would have been impossible without the expertise of Simone Schrott). Second, the copyright protection accorded to individual text contributions, such as reviews, synopses, etc, elided the attention of all communities we examined, and even communities, where pains were taken to properly indicate the licenses of the various parts of the database, existing incompatibility issues were not necessarily registered.

[18]

Finally, turning to the solution we came up with in the JVMG project.

[19]

The solution had to address the concerns of the communities we are working with.

[20]

As well as the long-term needs of the project itself. Not only did we have to decide on a lowest common denominator that would allow for the largest possible datasets to be integrated in the future, but also had to deal with the fact that both our source databases and our potential end-users reside in a range of different jurisdictions.

[21]

We settled on the CC BY-NC-SA (attribution, non-commercial, share-alike) 4.0 license, as it explicitly covers databases, is ported internationally, and because it is usually easier to ask communities with more permissive licenses for a separate non-commercial license for our project. We obtained this license in separate agreements for the project from the communities who had no license in place or whose licenses would not have been compatible with this CC license. Importantly, these separate license agreements only cover the parts of the community databases used in our project database, thereby mitigating some of the concerns that communities had about the potential wholesale copying of their databases. Finally, we omitted all database elements, such as reviews, summaries, etc, that would fall under the individual copyright claims of their respective authors.

[22]

This final table offers an overview of the ways license compatibility with the CC BY-NC-SA 4.0 license was reached for the various databases that the project database builds on.

[23]

Thank you very much for your attention. Please get in touch with us or visit our project website and the JVMG database.