DMLawTool – A Guiding Tool for Researchers to Address Legal Aspects in Data Management

Anna Picco-Schwendener
eLearning Lab
USI – Università della Svizzera italiana
Lugano, Switzerland
anna.picco.schwendener@usi.ch

Branislava Trifkovic
eLearning Lab
USI-Università della Svizzera italiana
Lugano, Switzerland
branislava.trifkovic@usi.ch

Suzanna Marazza
eLearning Lab
USI – Università della Svizzera italiana
Lugano, Switzerland
suzanna.marazza@usi.ch

***Abstract*—*DMLawTool is a web-based tool that guides researchers working mainly in the fields of humanities and social sciences through the most relevant legal issues related to data management. It has the form of decision tree that provides different solution approaches on how to correctly deal with research data during the whole data lifecycle. The released version of the DMLawTool will be available openly and free of charge as an open-source software on the CCdigitallaw.ch platform (www.ccdigitallaw.ch). This paper illustrates the different development steps of the tool, introduces its structure and main functionalities and provides a reflection on the faced challenges.*

Keywords*— *research data management, data protection, copyright, legal aspects;

I. INTRODUCTION

DMLawTool is currently being developed by the Università della Svizzera italiana in collaboration with the University of Neuchatel within the P-5 program of swissuniversities (swissuniversities, 2021).

The DMLawTool helps researchers to understand which legal issues they have to consider when dealing with research data. The two main topics it addresses are data protection and copyright. With the help of a decision tree, the tool guides the user through the most relevant legal aspects and finally provides suggestions on how to correctly handle research data throughout its whole lifecycle, from data collection to data archiving or erasure (elimination of data).

The idea to develop such a tool originates from the need of research data archives to address also legal aspects when archiving data. To date research data archiving platforms in the fields of humanities and social sciences in Switzerland mainly focused on the development of technically safe and secure solutions. However, with the introduction of the European General Data Protection Regulation (GDPR) (European Parliament and Council of European Union, 2016) the need of having complete Data Management Plans (DPMs) for projects of the Swiss National Science Foundation (SNF) (Swiss National Science Foundation, 2021), the upcoming revision of the Swiss Federal Act on Data Protection (Schweizer Eidgenossenschaft, 2019) and the recently revised Swiss Federal Act on Copyrights and Related Rights (CopA) (Schweizer Eidgenossenschaft, 2020), it has become essential to address legal aspects of data management in order to act in a legally compliant way. However, there is still much incertitude around legal aspects. Even though the legal liability of research data lies with the researcher - also during the archiving process - platforms understood the importance of raising awareness about legal aspects and providing researchers guidance and support.

The project thus closely collaborated with five Swiss research data archiving platforms in the fields of social sciences and humanities, in order to understand their needs and get an overview of the legal problems they have already come across.

The DMLawTool wants to act as a light bulb that turns on whenever researchers feel lost within the misty legal jungle and are unsure about the best approach to adopt in order to handle and archive research data in a legally compliant way. It addresses questions ranging from copyright (e.g. Is my research data considered a work and thus protected by copyright? Do I own the rights to make my data available for re-use?) to data protection (e.g. Does my data contain personal data? If yes, do I really need to anonymize it? How do I correctly anonymize it? What do I have to consider when working with personal data?), and licensing (e.g. which licenses should I use if I want to make data available for re-use in an open way?). The tool encourages open access practices and re-use of data wherever possible. Furthermore, it pays particular attention to the vulgarization of the legal language.

The tool is currently developed in English language and will be available as open-source software on the CCdigitallaw.ch platform (www.ccdigitallaw.ch) by the end of March 2021. In this way all platforms dealing with humanities and social science research data in Switzerland can use it either as a standalone instrument or integrate it in their archiving processes.

The following chapters will present the empirical basis of the project, introduce the DMLawTool and its functionalities, and provide a short reflection on the main challenges faced during the project.

II. Empirical Basis

A. *Data Collection*

In order to learn about how research data archiving platforms currently deal with legal aspects, and to identify legal issues and questions that they have already come across, the project partners closely collaborated with five Swiss repositories active in the fields of social sciences and humanities: 1) DaSCH (https://dasch.swiss), a platform for humanities research data 2) Yareta (https://yareta.unige.ch), the research data repository of Geneva's Higher Education Institutions, 3) FORSbase (https://forsbase.unil.ch), a platform that provides access to data about social sciences studies in Switzerland, 4) Dodis (https://www.dodis.ch), the archive for diplomatic documents of Switzerland, and 5) Historisches Lexikon der Schweiz (https://hls-dhs-dss.ch), an encyclopedia made out of research data.

Between January and May 2020, the project team conducted a total of six interviews with representatives of these platforms (three in person and three online).

The interviews revealed the central role of the researchers in the archiving process and also the fact that the legal liability of archived data actually lies with the researchers. In light of this, the project team decided to interview also three researchers active in the fields of humanities and a director of a university's research support center. This allowed to consider also the needs of researchers and considerably enriched the insights gained from the collected data.

B. *Data Analysis & Findings*

All information gathered during the interviews has been systematically organized in an Excel Table with a particular attention towards legal issues that have already emerged, solution approaches applied so far, and expectations towards the DMLawTool.

Subsequently, visual maps have been created to group, structure and link contents within single thematic clusters such as copyright, data protection and expectations towards the tool. Figure 1 shows the thematic map with clusters related to copyright. The map shows the importance of concepts such as re-use of data, licenses, image rights and ownership of copyrights. With regard to data protection, anonymization and the use of personal data resulted to be particularly hot topics. As far as the tool is concerned, interviewees expect to see practical examples of do's and don'ts. They would further appreciate templates for different kind of contracts and consent forms. They like the idea of a decision tree with a simple and straightforward structure that guides the user. Last but not least they suggest using a language which is accessible to a public at large and including some interactive elements as well as references and

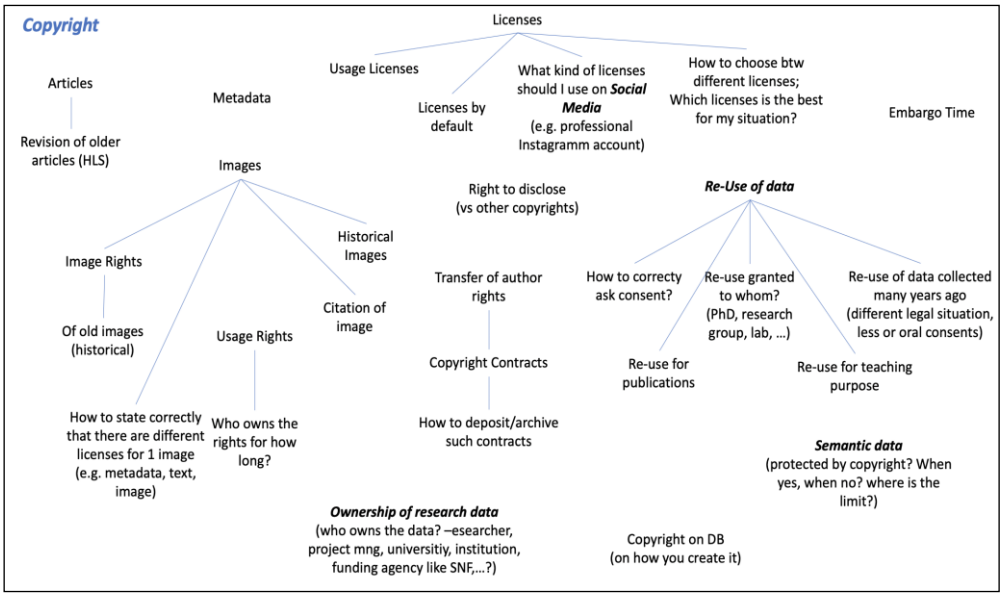

Fig. 1. Thematic map with cluster related to copyright

links to additional resources allowing to 1) further investigate certain aspects and 2) contact people who can provide further help.

III. DMLawTool and its Functionalities

This paper describes a beta version of the DMLawTool, which can be accessed at the following link: https://dmlawtool.web.app. It corresponds to the development stage of the tool at the end of January 2021 and might be subject to further changes.

A. *Decision Tree Structure and Behavior*

The DMLawTool is structured as a decision tree and as such is composed of branches and nodes.

When fully expanded, it provides an overview of the main legal issues related to data management, however, its main goal is to guide users through the various legal topics by providing the necessary explanations to move through the tree. This allows users to understand which legal aspects are relevant for their specific cases and to further deepen them. This is also the reason why the tree at first is not shown fully expanded but has to be opened-up step by step by the user. As shown in Figure 2, a "+" sign on the node allows to expand the next branches while a "-" sign allows to collapse nodes.

Each node has a meaningful name and is connected to two or more child nodes. Users are asked to choose among various available branches to proceed within the tree. At the end of each branch there is a so-called "end-node". These nodes are highlighted with a different color (currently yellow) and generally represent specific actions that the user can or should take once arrived at this point. In other words, the end-nodes provide solutions to the legal issues identified as relevant for a specific situation.

Beyond the name, each node contains knowledge notions that are necessary to advance through the tree. They can be accessed by clicking on the node. The node texts have a precise guiding role, and do not follow the rules of a traditional encyclopedia, which would simply explain the term of the node. In fact, each node text provides those explanations and definitions necessary to understand the following nodes, so that a user is able to decide which branch to choose next. This means, that you will not find a definition of data protection in the data protection node, but you will find this definition in the previous node, as you need to know what data protection is in order to decide whether it is relevant for your current research.

Each node text is structured in a similar way. Even though the tool is about legal issues, the storytelling is

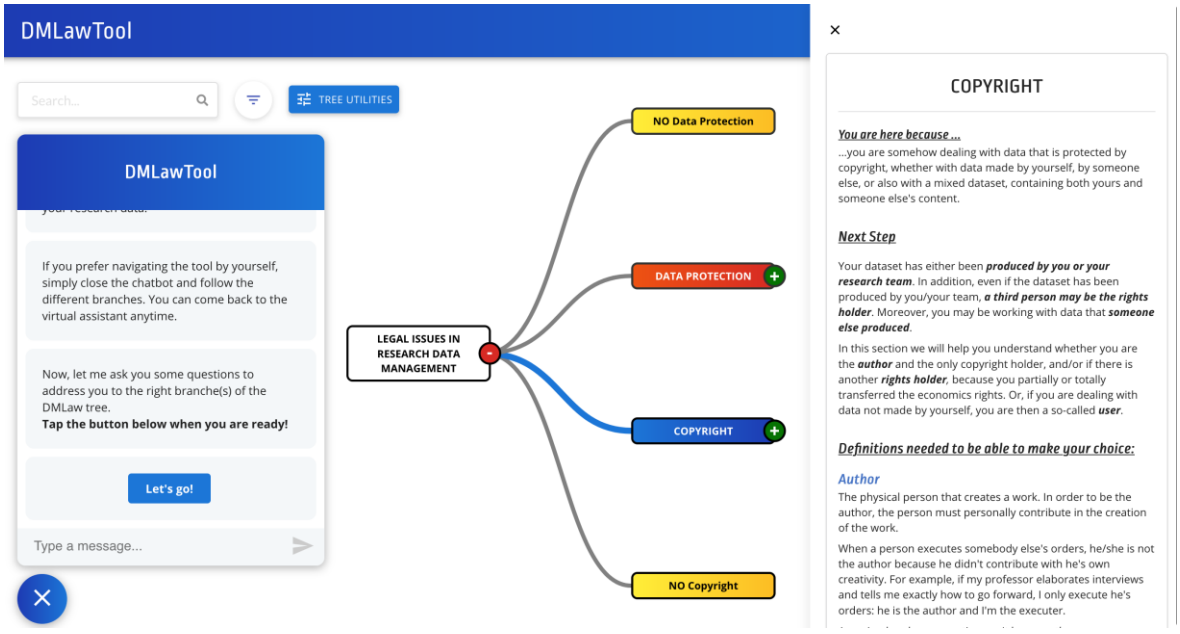

Fig. 2. DMLawTool with 1st branches of tree, text of copyright node and Chabot

fundamental. This is why each node text starts with a section called "You are here because …" followed by "Your next steps". Then the necessary definitions are provided and illustrated with practical examples. Each text ends with a section called "be aware of" and FAQs. End-nodes have a slightly different structure as they generally explain how to implement a specific action. For example, the end-node "anonymization" will illustrate different anonymization techniques.

Content wise, the decision tree is structured around data protection and copyright. Already at the first level there are also two end-nodes: “no data protection” and “no copyright”. These nodes show the potentials and pitfalls for your research data, if copyright or data protection regulations do not apply. In these cases, in fact, archiving, sharing and making data available as open access and for re-use is particularly easy as there are not many obstacles. These nodes thus allow favoring open access practices.

B. Chatbot

Another important element of the DMLawTool is the interactive Chabot, which simulates a conversation with the users. Its main function is to welcome users to the DMLawTool and introduce the decision tree, its behavior and functionalities. At first, it has been developed to lead the users to the first branches of the tree, however, after some user testing it emerged that in this way contents are provided in two different locations, which might be confusing. However, the designers are still thinking about other ways to take advantage of the Chabot in order to maximize its potential. The Chabot can be closed and re-opened at any time.

C. Search Functionalities and Tree Utilities

DMLawTool has a powerful text search function. When you enter a word, the tool highlights all nodes of the tree in which the word is present. Furthermore, each node has a series of tags assigned. This allows filtering contents based on chosen tags.

Tree Utilities allow to center and expand the tree and to zoom in and out. Furthermore, the tool works a lot with colors. Each of the two main legal topics have a dedicated color: data protection nodes are red while copyright nodes are blue. This allows to visually highlighting to which thematic area a node belongs.

IV. Main Challenges

A. Selecting the most Relevant Legal Aspect

DMLawTool focuses on legal aspects that are relevant for research data management. This means that the tool is a tradeoff between explaining every single legal detail or exception and providing best practices that are relevant for the people working with research data. The process of identifying aspects that could be neglected was not always easy. To guarantee a high quality of the tool, each item in question has thus been discussed extensively among the different legal experts involved in the project.

B. Vulgarization of legal language and storytelling

It is known that legal people speak their own language, which is often incomprehensible to non-experts. Using a simple vocabulary and creating a clear narrative, were thus among the main goals of the project. In order to find a balance between legal accuracy and general comprehension, legal and communication experts worked hand in hand: legal experts wrote the definitions while the communication experts elaborated the storytelling and included/vulgarized the definitions. In this way, all texts went through various iterations in order to be checked for their solidity from both a legal and communication point of view.

Another challenge was to avoid content duplications. Each legal aspect should have its specific place within the decision tree structure. This is important for the tool’s maintenance as it should be possible to implement changes in only one single place.

V. Conclusion

With the help of a carefully developed decision tree, the DMLawTool helps researchers to identify those legal aspects that are relevant for their projects and proposes them different solutions on how to handle their data in a legally compliant way. In this way, the tool wants to take away fears and favor open access, sharing and re-use of research data.

The focus of the tool has been widened during the project to cover not only the archiving stage but the whole life cycle of research data, from data collection to its archiving or cancellation.

Scalability is a very important aspect of the tool in order to allow adaptations or extensions. The tool has been developed in English, but translations in other languages are welcome and easily possible thanks to the fact that all contents will be published under an open Creative Commons license (Creative Commons, n.d.) and the tool will be available as open-source software.

Acknowledgment

The DMLawTool project is supported by the P-5 program of swissuniversities, under grant number 192-009.

We would like to particularly thank all the interviewees for their availability and insightful contributions in the data collection phase.

A warm thanks goes to all people involved in the project, in particular to Nathalie Tissot for the scientific supervision, Ivan Pavic for the software implementation, Yves Bauer for his legal inputs, and Mattia Pera for the graphic design.

References

Creative Commons. *Share your work*. Creative Commons. https://creativecommons.org/share-your-work/

European Parliament and Council of European Union. (2016). *Regulation (EU) 2016/679*. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN.

Schweizer Eidgenossenschaft. (01.03.2019). *Federal Act on Data Protection.* Fedlex. https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/en

Schweizer Eidgenossenschaft. (01.04.2020). *Federal Act on Copyright and Related Rights*. Fedlex. https://www.fedlex.admin.ch/eli/cc/1993/1798_1798_1798/en

Swiss National Science Foundation. (2021). *Data Management Plan (DMP) - Guidelines for researchers*. SNF. http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx

Swissuniversities. (2021). *P-5: Scientific information*. Swissuniversities. https://www.swissuniversities.ch/en/topics/digitalisation/p-5-scientific-information