

**Nobody puts  
research in a cage.**

**Researchers' perspectives  
on working with copyright.**

# Introduction

Access to Knowledge is key to the fundamental **Right to Research**. Resources used in the context of scientific research are often protected by copyright and related rights, and right holders can prohibit their use for research purposes. Researchers rely on copyright exceptions and limitations to **access, use and reuse** protected data sources in scientific projects. A fair and modern copyright framework is therefore essential to create an enabling environment for scientific research.

In some countries, researchers benefit from broad and flexible copyright exceptions and limitations that allow them to use protected materials in their projects, while in others they face overly restrictive laws that force them to either refrain from using such materials or to work in legal grey zones.

In the European Union (EU), a recent reform attempted to address some of the obstacles copyright law poses to scientific research. Text and data mining – a modern technology where researchers use computational methods to analyse mass amounts of text, images and other data

sources – is now allowed across all the EU Member States. While the EU-wide copyright exception for text and data mining represents a significant improvement to the legal framework for research in the region, it does not respond to all the pressing needs of researchers and their audiences. The mandatory exception only covers the rights of reproduction. It does not cover the right of communication to the public, which is essential to enable researchers to access research resources remotely and to share the research results and underlying resources for purposes of verification, validation and dissemination of results. This poses problems from the perspective of **research transparency**, prevents researchers from complying with open access requirements for scientific research and hinders **joint and cross-border** initiatives.

This publication intends to demonstrate some of the problems that EU researchers face due to these constraints. It assembles an initial selection of views of individual researchers from Sweden, Poland, Italy and Slovenia that were gathered through a series of interviews conducted in January

and February 2023 as part of the Right to Research in International Copyright Law project. The aim of the initiative is to better understand the needs and challenges faced by European researchers, particularly those interested in joint research activities and cross-border research collaborations. The interviews discuss issues of access, use and reuse of knowledge in unilateral and multilateral projects and the interviewees' perception of the current limitations related to copyright having an impact on their work.

So far, twenty interviews were conducted, with more data gathering planned for the near future, including more perspectives from female researchers. The final overview will be openly shared in the course of 2023.

The presented views do not constitute a representative research sample. Nevertheless, they serve as strong evidence supporting the need for a **global copyright reform** in the field of research addressing the essentials outlined by the researchers.

# Interviews

## Introducing the researchers



**Anamaria  
Dutceac Segesten**

I am a Senior Lecturer in European Studies at the Center for Language and Literature at Lund University (Lund, Sweden). I am also a reader at the department of strategic communication at the same university. Right now I am doing research on social media and politics, or on a more abstract level about the intersection between technology and democracy.



**Angelo Mario  
Del Grosso**

I am a researcher (level III) at CNR-ILC (Consiglio Nazionale delle Ricerche, Istituto di Linguistica Computazionale) (Pisa, Italy). My research interests span from designing and implementing applications for Digital Humanities to build digital textual corpora for philological studies.



**Jonas  
Ingvarsson**

I am a Senior Lecturer in Comparative Literature at the University of Gothenburg (Gothenburg, Sweden). I am also responsible for the master's program in digital humanities at the University of Gothenburg. I did my dissertation on literature and technology / cybernetics in the Swedish 1960s and have since been working on media history and literature.



**Måns  
Magnusson**

I am an Assistant Professor in Statistics at Uppsala University (Uppsala, Sweden). I am active in three to four different research fields, where we try to draw statistical conclusions from large amounts of textual data. I do this in collaboration with researchers in history, political science, law, sociology, and so on.



**Maciej  
Maryl**

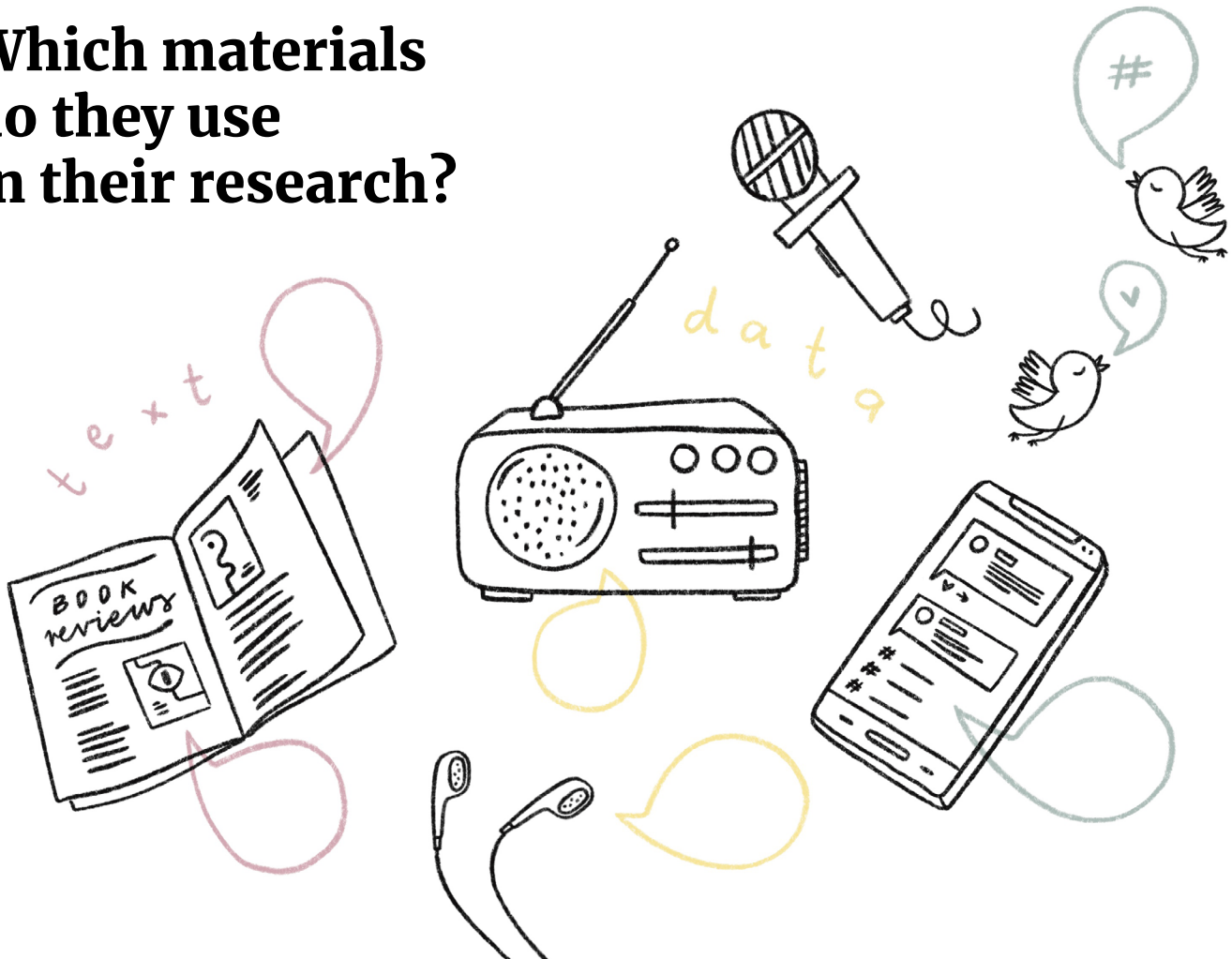
I am the Director of the Digital Humanities Center at the Institute of Literary Research of the Polish Academy of Sciences (Warsaw, Poland). As a researcher, I am involved in analysing bibliographic data, data-driven sociology of literature and research infrastructures.



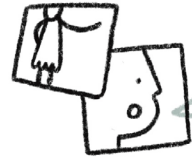
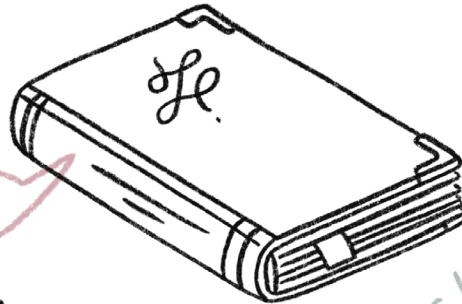
**Mitar  
Milutinović**

I am a researcher at Layer8 Institute (Slovenia). I was an AI researcher at UC Berkeley. My background is in computer science and my main current area of research is artificial intelligence, more precisely automated machine learning.

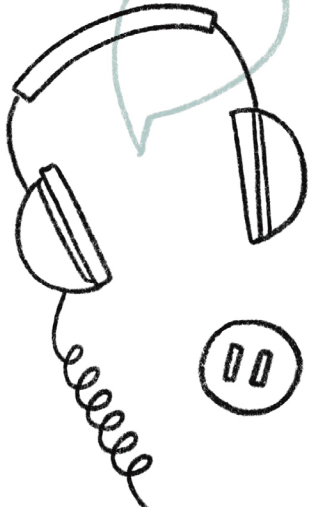
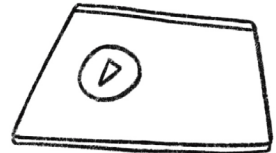
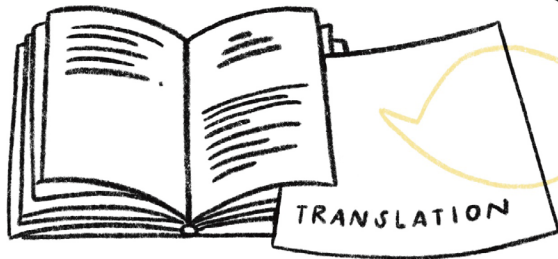
# Which materials do they use in their research?



copyrighted



machine learning



In their research Anamaria, Angelo, Jonas, Maciej, Måns and Mitar rely heavily on text and data mining and machine learning. They work with different types of copyrighted material: newspapers, novels, literary magazines, translations, social media data, radio broadcasts, datasets with images and videos.

# How are they using those materials?



**Måns:** We research public discourse and analyse daily newspapers, radio broadcasts, etc. We really get into copyright issues and there it has been very messy. Much of it is based on the kind of research that is becoming commonplace right now in Europe and the US, ie. text as data – where text is treated as data for scientific research.



**Anamaria:** I am investigating the media discourse around refugee flows over a five-year period, to examine the evolution of the migration discourse over time (including the Syrian refugee crisis).



**Maciej:** In one of my projects, we have created a corpus of literary discourse for the past 200 years in Poland. We are trying to get copyright permissions to get the data from publishers.

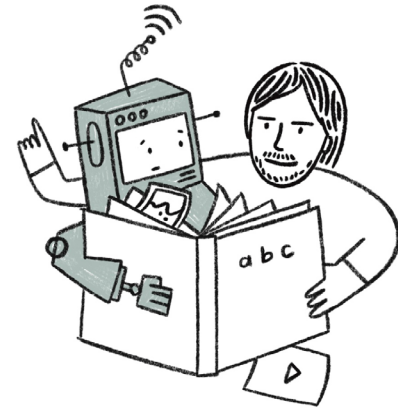




**Jonas:** We are studying book reviews in Swedish newspapers from 1906, 1956 and 2006. We want to train the computers to understand different expressions in their context. We also have a dream that feels more and more likely, insane at first but now maybe real? That is, to train a text corpus to identify what is a book review!



**Angelo:** I use copyrighted materials to extract info and recognize patterns and build indices to perform textual retrieval tasks. In my research project, we need to collect bilingual scientific publications (papers, monographs, metadata) in order to create bilingual corpora for machine translation training, terminology extraction and translation memory creation.



**Mitar:** We have made AI models which were trained on code and data to be able to process that data automatically.

# What copyright-related challenges do they face?

**Maciej:** The first obstacle would be access to texts and data. The second one would be the shady border between copyright and fair use, so having regulations that are understandable by non-legals. The third one is on how to licence your output and share it legally. When it comes to literary corpora, if you collect texts to do machine learning to support your research, it is really difficult to share them later or even to access them.

The fourth obstacle would be the lack of proper legal guidance - even if you have an institutional lawyer and legal department, they are not always trained in this particular issue. And they are conservative by default as they are trying to protect the interest of the institution.

**Måns:** Our big problem at the moment is that because this type of material is protected by copyright, it can only be analysed at the National Library [in Stockholm]. The research community is really annoyed at this, it's difficult for everyone outside of Stockholm.

Many projects that I would like to do, I have not even considered, I have just assumed that it is not possible. If I can't analyse the Swedish daily press in Umeå, I can't do it in Finland.

If I want to research French material, I must collaborate with French people and assume that they are aware of their copyright. I would never have dared to do research on Le Monde or anything like that without collaborating with someone in France.

**Jonas:** To access material from 1956, we have to go to the National Library Lab in Stockholm. It is a small glass cage with three data terminals. You sit in the lab, annotate. Access to it costs SEK 70,000 the first year, and 35,000 in the following years. You are not allowed to take data in or out, all labs must be done in the cage.

The transparency is non-existent. If someone wants to verify the results, they also have to buy the license for a lot of money. An incredible anxiety!





**Anamaria:** I purchased data from Twitter, Reddit and Youtube. The agreements oblige me not to share data except within the project. The data must be on my server. Only researchers in the consortium have access to my data. We are not allowed to share the data for peer review, which is a huge problem.

Perhaps there are situations where, for example, Swedish researchers cannot share data from, for example, an experiment, but Norwegian researchers can, and then either the Swedish researchers break the law or they cannot publish in the journals that require open data.

In the future, we will see that social media (web data, chat apps, etc.) will be used less in research because there are so many requirements and it is so cumbersome. People will say it's not worth it. There must be a better balance, in legislation and structures, that makes exceptions and establishes means and methods for those who have legitimate purposes to study something that is problematic.

**Angelo:** In my experience, a lot of constraints have been raised from copyright laws both in using and in publishing digital resources. Mainly I have to deal with publishing companies and personal successors of cultural material.

The main copyright-related obstacles concern awareness about licence topics when new Digital Humanities projects start. Moreover, copyright owners are unaware of the constraints and consequences of the licences to which they have subscribed. In general, there is a lack of law awareness.

**Mitar:** The main limitation we encountered was that we wanted to train our AI models on academic papers describing novel AI models (our project uses AI to build new AI models themselves, so training it on human AI models helps), but many papers are behind paywalls so we were not able to access them.

Another issue was collecting various datasets which exist on the web, that often lack clear information about licensing/copyright of the data so it is unclear if you can use them, even if you can access them.

And finally, after we have collected all this data, we wanted to make it public, but for many datasets we were unable to confirm that we have redistribution rights so we were unable to do so.

**Do they value  
collaborating  
across borders?**



**Angelo:** The main value of supra-national joint research initiatives is exchanging knowledge, best practices, and opportunities to open new research questions from different perspectives and experiences. Cross-national networking is crucial for our research field. Colleagues and resources are spread around the world. European and pan-European research infrastructures are particularly important to guarantee sustainability of DH initiatives.

**Anamaria:** What gives me the greatest joy is being able to collaborate with people who are doing similar things elsewhere. In my department and at Lund University, there are not many others who do the same thing as me, or who can relate their research to my research – even though it is a large university. I imagine that the problem becomes even more urgent at smaller universities. Complementary to that is finding people who do other things that I can't do here locally. They can bring their perspectives and their methodological skills, creating a complementary project. I have many such collaborations, which result in publications but also in other initiatives supporting fellow researchers, students, new networks and so on.

**Måns:** I collaborate with researchers in many different countries, such as the UK, Germany and so on. Many of the scientific questions we have are not national. Such as in my research project: what effect do different events have on the migration discourse. If you can only analyse it in Sweden, you get a small piece of the puzzle, the more countries you can look at, the more puzzle pieces you can find, and thus add more puzzle pieces to the big knowledge puzzle.

**Maciej:** I am doing my research in Poland and internationally, mostly in cooperation with European colleagues. In the case of bibliographic data, it is fascinating that we can work not only on Polish culture and what is published in Poland but also through international collaborations we can investigate which authors are translated and why etc. Another big goal is that we can exchange methodologies and ideas on how to do things. In my research I am dealing with a niche thing, working on specific data in Poland. However, in other countries there might already be very experienced researchers and some methodologies I can use or get in dialogue with them. The third value is to get the critical mass when working on something.

# Recommendations

## **Protect the right to research at the national level**

National policy makers should amend their copyright laws to permit research uses. Researchers should be allowed to conduct research on all kinds of copyrighted materials. They should be able to reproduce protected materials as well as to share them, in order to facilitate collaborative research and research transparency. It is equally important to ensure that research uses are protected from contractual and technological overrides.



## **Protect the right to research at the EU level**

Text and data mining is already mandatory in the European Union. The EU now needs to enable other types of research uses as well as joint research initiatives, including across borders. This requires an EU-wide mandatory research exception to copyright and other exclusive rights that permits, among other uses, sharing of protected materials.

## **Protect the right to research at the international level**

Policy makers should develop non-binding instruments on research uses, to support countries when reforming copyright laws to create an enabling environment for scientific research. In order to ensure that research is permitted in all countries and to foster research collaboration across borders, policy makers should also reach an international agreement on a set of minimum standards for research.

**Editors:**

Teresa Nobre, COMMUNIA  
Maria Drabczyk, Centrum Cyfrowe  
Eric Luth, Wikimedia Sverige  
Deborah De Angelis, DDA Law Firm  
Maja Bogataj Jančič, Intellectual Property Institute

**Designer:**

Magda Araźny, magdarysuje.pl

**Acknowledgements:**

We would like to express our gratitude to all the individuals who accepted our invitation to participate in the interviews that supported this publication.



COMMUNIA, Centrum Cyfrowe, Wikimedia Sverige and the Intellectual Property Institute are members of the Access to Knowledge Coalition – a group of civil society organisations from all around the globe that promotes access to, and use of, knowledge. Learn more about the Right to Research at [a2k-coalition.org](https://a2k-coalition.org).

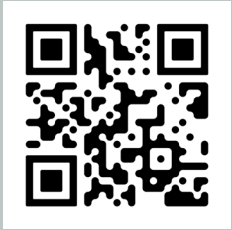


The publication is issued as part of the Right to Research in International Copyright Law project funded by Arcadia – a charitable fund of Lisbet Rausing and Peter Baldwin.



To the extent possible under law, all parties involved have waived all copyright and related or neighbouring rights to this publication (but not to the individual drawings). You can copy, modify, distribute, communicate to the public, even for commercial purposes, all without asking permission

March 2023



The QR code will take you  
to the online version  
of the publication.