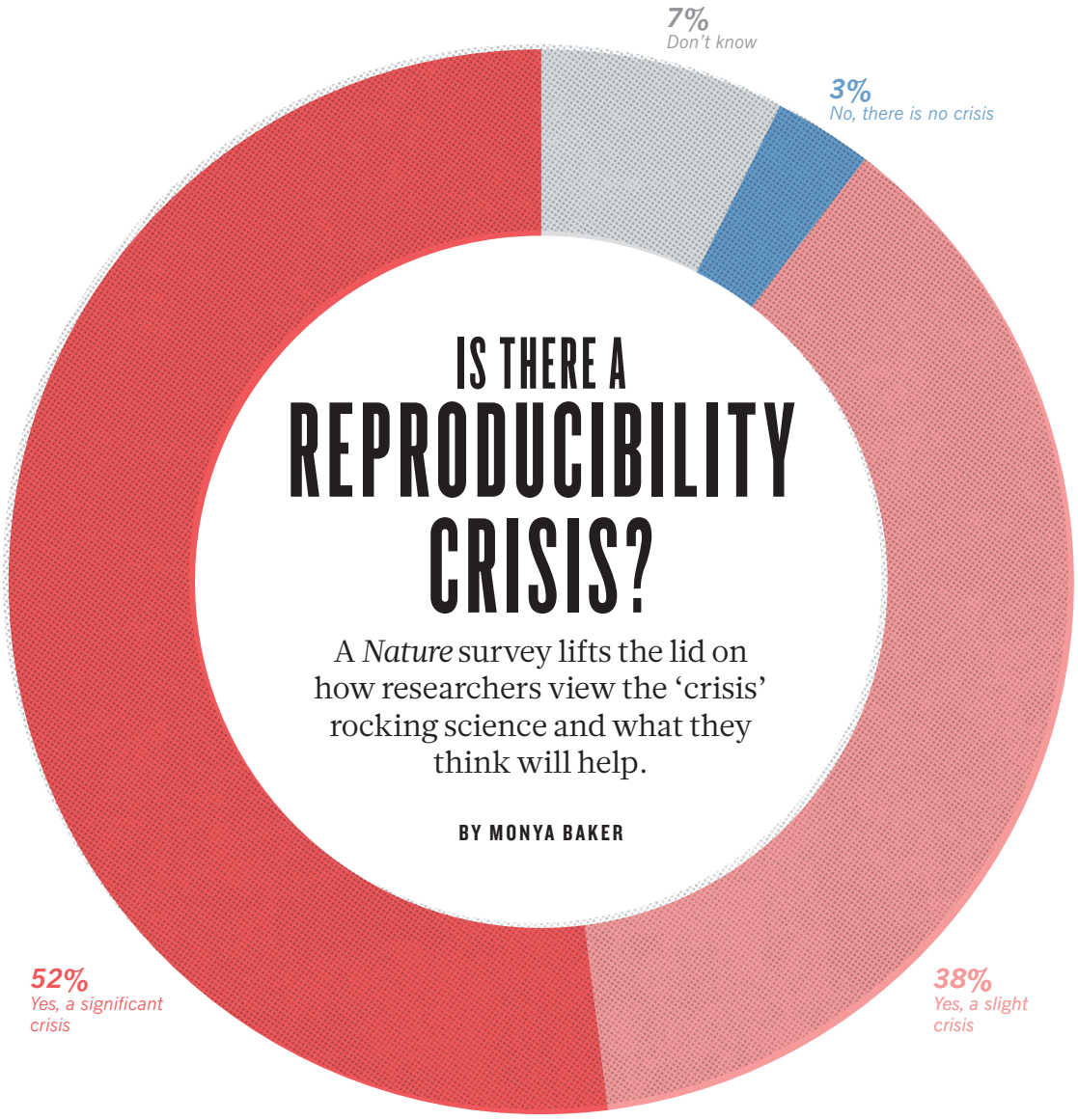


IS THERE A REPRODUCIBILITY CRISIS?

A *Nature* survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help.

BY MONYA BAKER

More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

The data reveal sometimes-contradictory attitudes towards reproducibility. Although 52% of those surveyed agree that there is a significant 'crisis' of reproducibility, less than 31% think that failure to reproduce published results means that the result is probably wrong, and most say that they still trust the published literature.

Data on how much of the scientific literature is reproducible are rare and generally bleak. The best-known analyses, from psychology[1] and cancer biology[2], found rates of around 40% and 10%, respectively. Our survey respondents were more optimistic: 73% said that they think that at least half of the papers in their field can be trusted, with physicists and chemists generally showing the most confidence.

The results capture a confusing snapshot of attitudes around these issues, says Arturo Casadevall, a microbiologist at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. "At the current time there is no consensus on what reproducibility is or should be." But just recognizing that is a step forward, he says. "The next step may be identifying what is the problem and to get a consensus."

Failing to reproduce results is a rite of passage, says Marcus Munafo, a biological psychologist at the University of Bristol, UK, who has a long-standing interest in scientific reproducibility. When he was a student, he says, "I tried to replicate what looked simple from the literature, and wasn't able to. Then I had a crisis of confidence, and then I learned that my experience wasn't uncommon."

The challenge is not to eliminate problems with reproducibility in published work. Being at the cutting edge of science means that sometimes results will not be robust, says Munafo. "We want to be discovering new things but not generating too many false leads."

THE SCALE OF REPRODUCIBILITY

But sorting discoveries from false leads can be discomfiting. Although the vast majority of researchers in our survey had failed to reproduce an experiment, less than 20% of respondents said that they had ever been contacted by another researcher unable to reproduce their work (see 'A 'crisis' in numbers'). Our results are strikingly similar to another online survey of nearly 900 members of the American Society for Cell Biology (see go.nature.com/kbzs2b). That may be because such conversations are difficult. If experimenters reach out to the original researchers for help, they risk appearing incompetent or accusatory, or revealing too much about their own projects.

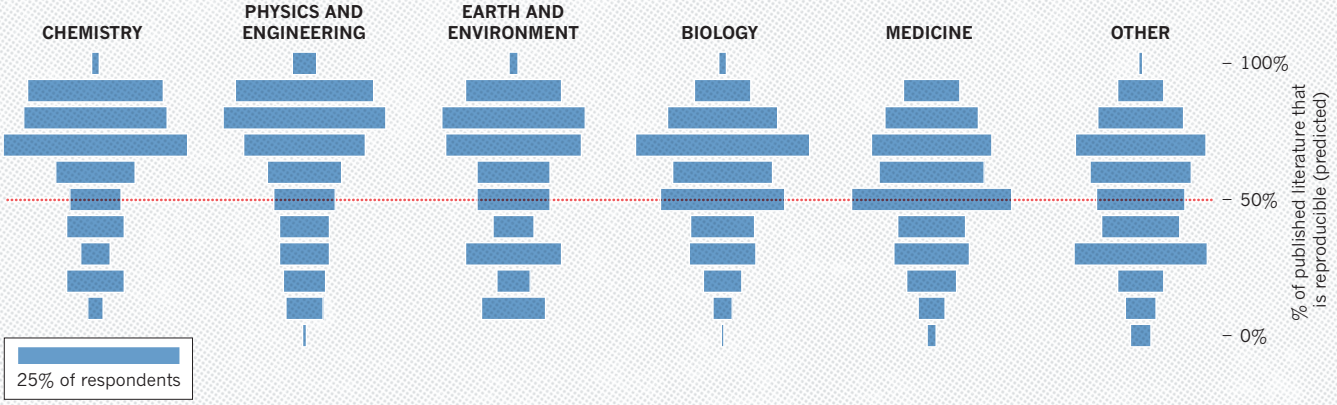
A minority of respondents reported ever having tried to publish

A 'CRISIS' IN NUMBERS

Nature surveyed 1,576 scientists online to get their thoughts on reproducibility in their field and in science in general. See **go.nature.com/2vjr4y** for more charts and access to the full data.

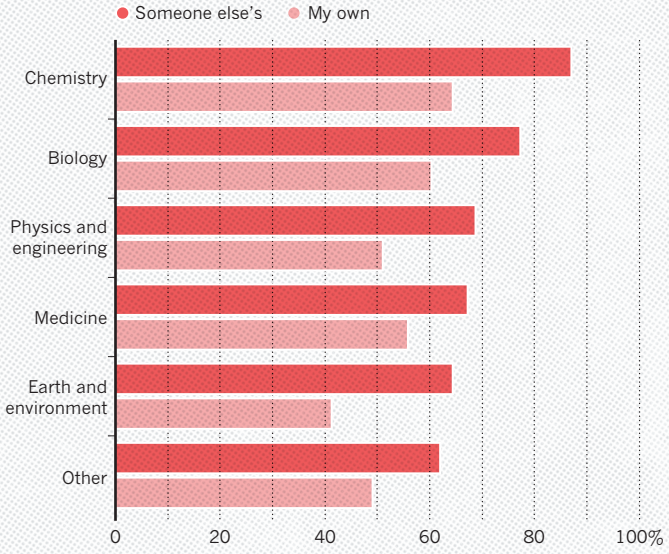
HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?

Physicists and chemists were most confident in the literature.



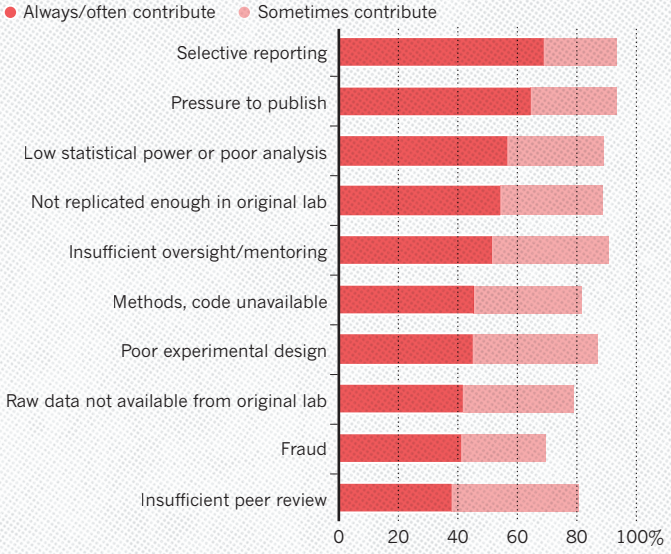
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



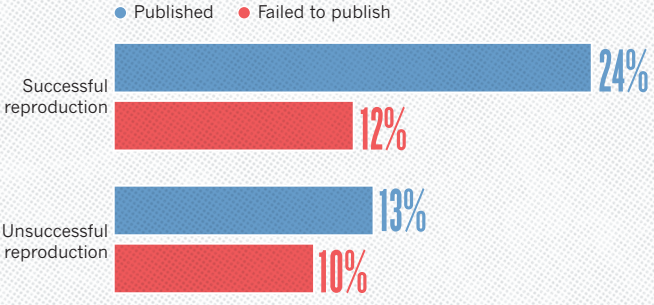
WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



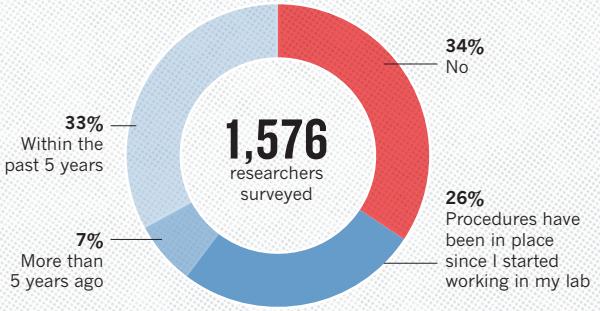
HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.



HAVE YOU ESTABLISHED PROCEDURES FOR REPRODUCIBILITY?

Among the most popular strategies was having different lab members redo experiments.



Number of respondents from each discipline:
Biology **703**, Chemistry **106**, Earth and environmental **95**, Medicine **203**, Physics and engineering **236**, Other **233**

a replication study. When work does not reproduce, researchers often assume there is a perfectly valid (and probably boring) reason. What's more, incentives to publish positive replications are low and journals can be reluctant to publish negative findings. In fact, several respondents who had published a failed replication said that editors and reviewers demanded that they play down comparisons with the original study.

Nevertheless, 24% said that they had been able to publish a successful replication and 13% had published a failed replication. Acceptance was more common than persistent rejection: only 12% reported being unable to publish successful attempts to reproduce others' work; 10% reported being unable to publish unsuccessful attempts.

Survey respondent Abraham Al-Ahmad at the Texas Tech University Health Sciences Center in Amarillo expected a "cold and dry rejection" when he submitted a manuscript explaining why a stem-cell technique had stopped working in his hands. He was pleasantly surprised when the paper was accepted[3]. The reason, he thinks, is because it offered a workaround for the problem.

Others place the ability to publish replication attempts down to a combination of luck, persistence and editors' inclinations. Survey respondent Michael Adams, a drug-development consultant, says that work showing severe flaws in an animal model of diabetes has been rejected six times, in part because it does not reveal a new drug target. By contrast, he says, work refuting the efficacy of a compound to treat Chagas disease was quickly accepted[4].

"REPRODUCIBILITY IS LIKE BRUSHING YOUR TEETH. ONCE YOU LEARN IT, IT BECOMES A HABIT."

THE CORRECTIVE MEASURES

One-third of respondents said that their labs had taken concrete steps to improve reproducibility within the past five years. Rates ranged from a high of 41% in medicine to a low of 24% in physics and engineering. Free-text responses suggested that redoing the work or asking someone else within a lab to repeat the work is the most common practice. Also common are efforts to beef up the documentation and standardization of experimental methods.

Any of these can be a major undertaking. A biochemistry graduate student in the United Kingdom, who asked not to be named, says that efforts to reproduce work for her lab's projects doubles the time and materials used — in addition to the time taken to troubleshoot when some things invariably don't work. Although replication does boost confidence in results, she says, the costs mean that she performs checks only for innovative projects or unexpected results.

Consolidating methods is a project unto itself, says Laura Shankman, a postdoc studying smooth muscle cells at the University of Virginia, Charlottesville. After several postdocs and graduate students left her lab within a short time, remaining members had trouble getting consistent results in their experiments. The lab decided to take some time off from new questions to repeat published work, and this revealed that lab protocols had gradually diverged. She thinks that the lab saved money overall by getting synchronized instead of troubleshooting failed experiments piecemeal, but that it was a long-term investment.

Irakli Loladze, a mathematical biologist at Bryan College of Health Sciences in Lincoln, Nebraska, estimates that efforts to ensure reproducibility can increase the time spent on a project by 30%, even for his theoretical work. He checks that all steps from raw data to the final figure can be retraced. But those tasks quickly become just part of the job. "Reproducibility is like brushing your teeth," he says. "It is good for you, but it takes time and effort. Once you learn it, it becomes a habit."

One of the best-publicized approaches to boosting reproducibility is pre-registration, where scientists submit hypotheses and plans for data analysis to a third party before performing experiments, to prevent cherry-picking statistically significant results later. Fewer than a dozen people mentioned this strategy. One who did was Hanne Watkins, a graduate student studying moral decision-making at the University of Melbourne in Australia. Going back to her original questions after collecting data, she says, kept her from going down a rabbit hole. And the process, although time consuming, was no more arduous than getting ethical approval or formatting survey questions. "If it's built in right from the start," she says, "it's just part of the routine of doing a study."

THE CAUSE

The survey asked scientists what led to problems in reproducibility. More than 60% of respondents said that each of two factors — pressure to publish and selective reporting — always or often contributed. More than half pointed to insufficient replication in the lab, poor oversight or low statistical power. A smaller proportion pointed to obstacles such as variability in reagents or the use of specialized techniques that are difficult to repeat.

But all these factors are exacerbated by common forces, says Judith Kimble, a developmental biologist at the University of Wisconsin–Madison: competition for grants and positions, and a growing burden of bureaucracy that takes away from time spent doing and designing research. "Everyone is stretched thinner these days," she says. And the cost extends beyond any particular research project. If graduate students train in labs where senior members have little time for their juniors, they may go on to establish their own labs without having a model of how training and mentoring should work. "They will go off and make it worse," Kimble says.

WHAT CAN BE DONE?

Respondents were asked to rate 11 different approaches to improving reproducibility in science, and all got ringing endorsements. Nearly 90% — more than 1,000 people — ticked "More robust experimental design" "better statistics" and "better mentorship". Those ranked higher than the option of providing incentives (such as funding or credit towards tenure) for reproducibility-enhancing practices. But even the lowest-ranked item — journal checklists — won a whopping 69% endorsement.

The survey — which was e-mailed to *Nature* readers and advertised on affiliated websites and social-media outlets as being 'about reproducibility' — probably selected for respondents who are more receptive to and aware of concerns about reproducibility. Nevertheless, the results suggest that journals, funders and research institutions that advance policies to address the issue would probably find cooperation, says John Ioannidis, who studies scientific robustness at Stanford University in California. "People would probably welcome such initiatives." About 80% of respondents thought that funders and publishers should do more to improve reproducibility.

"It's healthy that people are aware of the issues and open to a range of straightforward ways to improve them," says Munafo. And given that these ideas are being widely discussed, even in mainstream media, tackling the initiative now may be crucial. "If we don't act on this, then the moment will pass, and people will get tired of being told that they need to do something." SEE EDITORIAL P.437

Monya Baker *writes and edits for* Nature *from San Francisco.*
Dan Penny *aided in creation and analysis of the survey.*

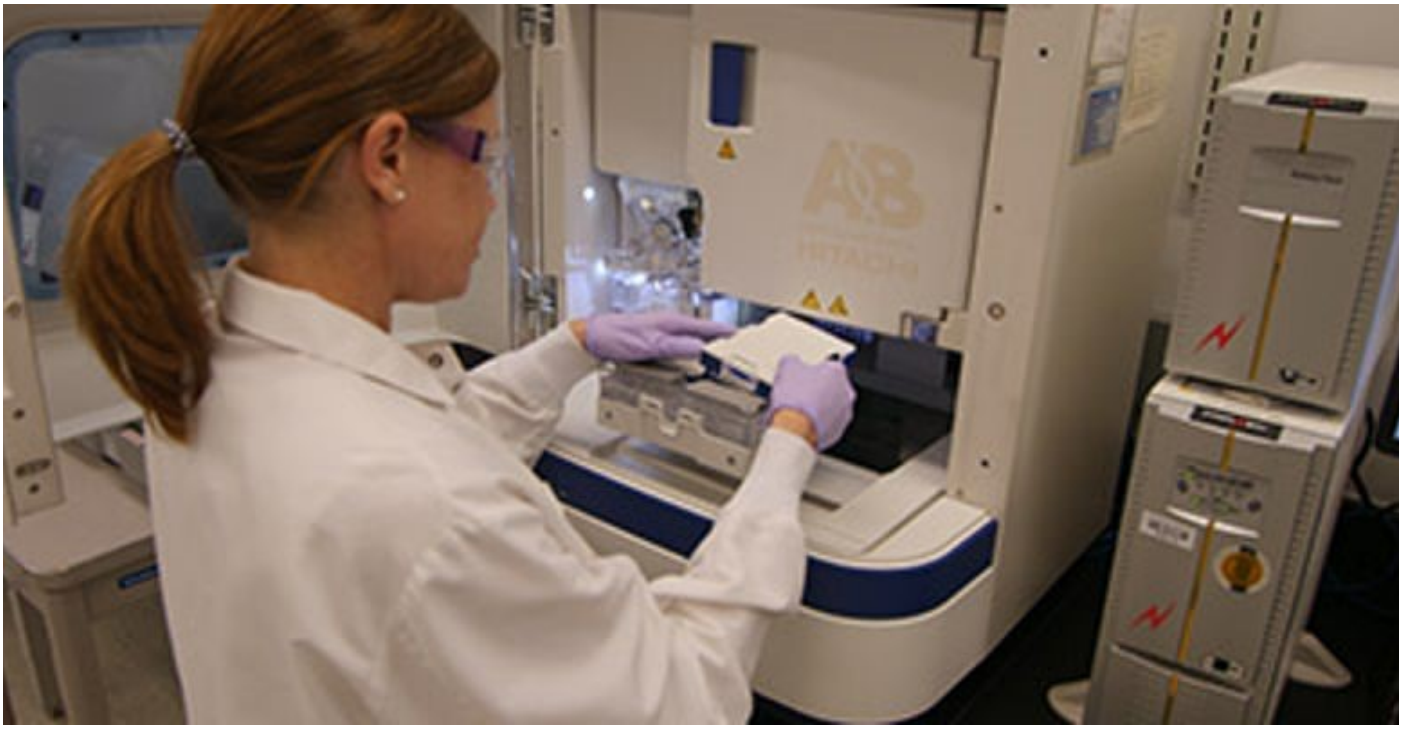
1. Open Science Collaboration *Science* http://dx.doi.org/10.1126/science.aac4716 (2015).
2. Begley, C. G. & Ellis, L. M. *Nature* **483,** 531–533 (2012).
3. Patel, R. & Alahmad, A. J. *Fluids Barriers CNS* **13,** 6 (2016).
4. da Silva, C. F. *et al. Antimicrob. Agents Chemother.* **57,** 5307–5314 (2013).

Six factors affecting reproducibility in life science research and how to handle them

Scientific advancement depends on a strong foundation of data credibility. However, scientific findings in biomedical research are not always reproducible. Meet the organizations that are promoting best practices and helping researchers perform the highest-quality science.

Produced by

ATCC®



There are several reasons why an experiment cannot be replicated.

Independent verification of data is a fundamental principle of scientific research across the disciplines. The self-correcting mechanisms of the scientific method depend on the ability of researchers to reproduce the findings of published studies in order to strengthen evidence and build upon existing work. Stanford University medical researcher, John Ioannidis, a prominent scholar on reproducibility in science, has pointed out that the importance of reproducibility does not have to do with ensuring the 'correctness' of results, but rather with ensuring the transparency of exactly what was done in a given line of research[1].

In theory, researchers should be able to re-create experiments, generate the same results, and arrive at the same conclusions, thus helping to validate and strengthen the original work. However, reality does not always meet these expectations. Too often, scientific findings in biomedical research cannot be reproduced[2]; consequently, resources and time are wasted, and the credibility of scientific findings are put at risk. Furthermore, despite recent heightened awareness, there remains a significant need to better educate students and research trainees about the lack of reproducibility in life science research and actions that can be taken to improve it. Here, we review predominant factors affecting reproducibility and outline efforts to improve the situation.

What is reproducibility?

The phrase 'lack of reproducibility' is understood in the scientific community, but it is a rather broad expression that incorporates several aspects. Though a standardized definition has not been fully established, the American Society for Cell Biology® (ASCB®) has attempted a multi-tiered approach to defining the term *reproducibility* by identifying the subtle differences in how the term is perceived throughout the scientific community.

ACSB[4] has discussed these differences with the following terms: *direct replication*, which are efforts to reproduce a previously observed result by using the same experimental design and conditions as the original study; *analytic replication*, which aims to reproduce a series of scientific findings through a reanalysis of the original data set; *systemic replication*, which is an attempt to reproduce a published finding under different experimental conditions (e.g., in a different culture system or animal model); and *conceptual replication*, where the validity of a phenomenon is evaluated using a different set of experimental conditions or methods.

It is generally thought that the improvement of direct replication and analytic replication is most readily addressed through training, policy modifications, and other interventions, while failures in systematic and conceptual replication are more difficult to connect to problems with how research was performed as there is more natural variability at play.

The reproducibility problem

Many studies claim a significant result, but their findings cannot be reproduced. This problem has attracted increased attention in recent years, with several studies providing evidence that research is often not reproducible. A 2016 *Nature* survey[3], for example, revealed that in the field of biology alone, over 70% of researchers were unable to reproduce the findings of other scientists and approximately 60% of researchers could not reproduce their own findings.

The lack of reproducibility in scientific research has negative impacts on health, lower scientific output efficiency, slower[6,7] scientific progress, wasted time and money, and erodes the public's trust in scientific

research. Though many of these problems are difficult to quantify, there have been attempts to calculate financial losses. A 2015 meta-analysis[5] of past studies regarding the cost of non-reproducible research estimated that $28 billion per year is spent on preclinical research that is not reproducible. Looking at avoidable waste in biomedical research on the whole, it is estimated that as much as 85% of expenditure may be wasted due to factors that similarly contribute to non-reproducible research such as inappropriate study design, failure to adequately address biases, non-publication of studies with disappointing results, and insufficient descriptions of interventions and methods.

Factors contributing to the lack of reproducibility

Failures of reproducibility cannot be traced to a single cause, but there are several categories of shortcomings that can explain many of the cases where research cannot be reproduced. Here are some of the most significant categories.

A lack of access to methodological details, raw data, and research materials.

For scientists to be able to reproduce published work, they must be able to access the original data, protocols, and key research materials. Without these, reproduction is greatly hindered and researchers are forced to reinvent the wheel as they attempt to repeat previous work. The mechanisms and systems for sharing raw unpublished data and research materials, such as data repositories and biorepositories, need to be made robust so that sharing is not an impediment to reproducibility.

Use of misidentified, cross-contaminated, or over-passaged cell lines and microorganisms.

Reproducibility can be complicated and/or invalidated by biological materials that cannot be traced back to their original source, are not thoroughly authenticated, or are not properly maintained. For example, if a cell line is not identified correctly, or is contaminated with mycoplasma or another cell type, results can be affected significantly and their likelihood of replication diminished. There are many cases of studies conducted with misidentified or cross-contaminated cell lines, so results rendered questionable, and conclusions drawn from them are potentially invalid[8]. Improper maintenance of biological materials via long-term serial passaging can also seriously affect genotype and phenotype, which can make reproducing data difficult. Several studies have demonstrated that serial passaging can lead to variations in gene expression, growth rate, spreading, and migration in cell lines[9,10]; and changes in physiology, virulence factor production, and antibiotic resistance in microorganisms[11,12,13].

Inability to manage complex datasets

Advancements in technology have enabled the generation of extensive, complex data sets; however, many researchers do not have the knowledge or tools needed for analyzing, interpreting and storing the data

correctly. Further, new technologies or methodologies may not yet have established or standardized protocols, so variations and biases can be easily introduced, which in turn can affect the ability to analytically replicate the data.

Poor research practices and experimental design

Among the findings from scholarly efforts examining non-reproducibility is that, in a significant portion of cases, the cause could be traced to poor practices in reporting research results, and poor experimental design[14,15]. Poorly designed studies without a core set of experimental parameters, whose methodology is not reported clearly, are less likely to be reproducible. If a study is designed without a thorough review of existing evidence, or if the efforts to minimize biases are insufficient, reproducibility becomes more problematic.

Cognitive bias

These refer to the ways that judgement and decision-making are affected by the individual subjective social context that each person builds around them. They are errors made in cognitive processes that are due to personal beliefs or perceptions. Researchers strive for impartiality and try to avoid cognitive bias, but it is often difficult to completely shut out the subtle, subconscious ways that cognitive bias can affect the conduct of research[16,17]. Scientists have identified dozens of different types of cognitive biases, including confirmation bias, selection bias, the bandwagon effect, cluster illusion, and reporting bias[17]. Confirmation bias is the unconscious act of interpreting new evidence in ways that confirm one's existing belief system or theories; this type of bias impacts how information is gathered, interpreted, and recalled. Selection bias sees researchers choose subjects or data for analysis that is not properly randomized; here, the sample obtained is not truly representative of the whole population. The bandwagon effect is the tendency to agree with a position too easily, without sufficient evaluation in order to maintain group harmony; this form of bias may lead to the acceptance of unproven ideas that have gained popularity. Cluster illusion is when patterns are perceived in a pool of random data in which no actual pattern exists; a bias based on the tendency of the brain to seek out patterns. Reporting bias is when study participants selectively reveal or suppress information in a study according to their own subconscious drivers; this form of bias may lead to underreporting of negative or undesirable experimental results.

A competitive culture that rewards novel findings and undervalues negative results

The academic research system encourages the rapid publication of novel results. Researchers are rewarded more for publishing novel findings, and not for publishing negative results (e.g., where a correlation was not found)[15]. Indeed, there are limited arenas for publishing negative results, which could hone researchers' efforts and avoid repeating work that may be difficult to replicate. Overall, reproducibility in research is hindered by under-reporting of studies that yield results deemed disappointing or insignificant. University

hiring and promotion criteria often emphasize publishing in high-impact journals and do not generally reward negative results. Also, a competitive environment for research grants may incentivize researchers to limit reporting of details learned through experience that make experiments work better.

Recommended best practices

A number of significant efforts have been aimed at addressing the lack of reproducibility in scientific research. Individual researchers, journal publishers, funding agencies, and universities have all made substantial efforts toward identifying potential policy changes aimed at improving reproducibility[16,18,19,20,21]. What has emerged from these efforts is a set of recommended practices and policy prescriptions that are expected to have a large impact.

Training on statistical methods and study design is essential for reproducible research.

Robust sharing of data, materials, software, and other tools.

All of the raw data that underlies any published conclusions should be readily available to fellow researchers and reviewers of the published article. Depositing the raw data in a publicly available database would reduce the likelihood that researchers would select only those results that support a prevailing attitude or confirms previous work. Such sharing would accelerate scientific discoveries, and enable scientists to interact and collaborate at a meaningful level.

Use of authenticated biomaterials

Data integrity and assay reproducibility can be greatly improved by using authenticated, low-passage reference materials. Cell lines and microorganisms verified by a multifaceted approach that confirms phenotypic and genotypic traits, and a lack of contaminants, are essential tools for research. By starting a set of experiments with traceable and authenticated reference materials, and routinely evaluating biomaterials throughout the research workflow, the resulting data will be more reliable, and more likely to be reproducible.

Training on statistical methods and study design

Experimental reproducibility could be considerably improved if researchers were trained how to properly structure experiments and perform statistical analyses of results. By strictly adhering to a set of best practices in statistical methodology and experimental design, researchers could boost the validity and reproducibility of their work.

Pre-registration of scientific studies

If scientists pre-register proposed scientific studies (including the approach) prior to initiation of the study, it would allow careful scrutiny of all parts of the research process and would discourage the suppression of negative results.

Publish negative data

Many times, ‘negative’ data that do not support a hypothesis typically go unpublished as they are not considered high impact or innovative. By publishing negative data, it helps to interpret positive results from related studies and can help researchers adjust their experimental design so that further resources and funding are not wasted[22].

Thorough description of methods

It is important that research methodology is thoroughly described to help improve reproducibility. Researchers should clearly report key experimental parameters, such as whether experiments were blinded, which standards and instruments were used, how many replicates were made, how the results were interpreted, how the statistical analysis was performed, how the randomization was done, and what criteria were used to include or exclude any data.

Ongoing efforts to improve reproducibility

There is a varied and influential group of organizations that are already working to improve the reproducibility of scientific research. The following is a list of initiatives aimed at supporting one or more

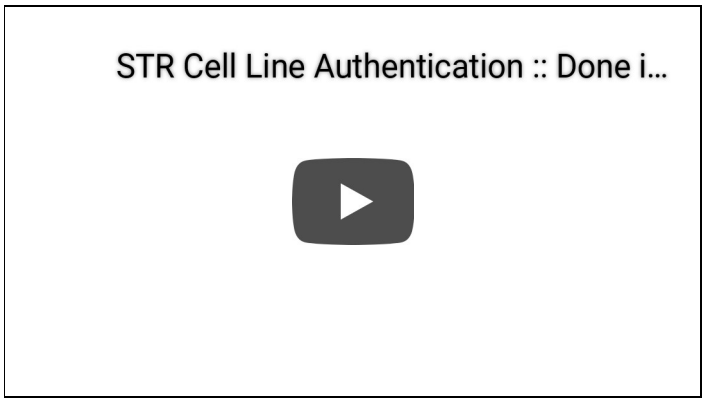
aspects of the research reproducibility issue.

American Society for Cell Biology (ASCB) - The ASCB Report on Reproducibility

ASCB continues to identify methods and best practices that would enhance reproducibility in basic research. From its original analysis, the ASCB task force identified and published several recommendations focused on supporting existing efforts and initiating new activities on better training, reducing competition, sharing data, improving peer reviews, and providing cell authentication guidelines.

American Type Culture Collection (ATCC) - Cell and Microbial Authentication Services and Programs

Biological resource centers, such as ATCC, provide the research community with standardized, traceable, fully authenticated cell lines and microorganisms to aid in assay reproducibility. At ATCC, microbial strains are authenticated and characterized through genotypic, phenotypic, and functional analyses to confirm identity, purity, virulence, and antibiotic resistance. ATCC has also taken a lead in cell line authentication by publishing the voluntary consensus standard, *ANSI/ATCC ASN-0002: Authentication of Human Cell Lines: Standardization of STR Profiling*, and by performing STR profiling on all human cell lines managed among its holdings.



Furthermore, ATCC offers online cell line authentication training in partnership with Global Biological Standards Institute, NIH (R25GM116155-03), and Susan G. Komen (SPP160007), which focuses on the best practices for receiving, managing, authenticating, culturing, and preserving cell cultures. To further support cell authentication and reproducibility in the life sciences, ATCC also provides STR profiling and mycoplasma detection testing as services to researchers.

National Institutes of Health (NIH) - Rigor and Reproducibility

To help improve rigor, reproducibility, and transparency in scientific research, the NIH issued a notice in 2015 that informed scientists of revised grant application instructions that focused on improving experimental design, authenticating biological and chemical resources, analyzing and interpreting results, and accurately reporting research findings. These efforts have led to the adoption of similar guidelines by journals across numerous scientific disciplines and has resulted in cell line authentication becoming a prerequisite for publication.

Science Exchange & the Center for Open Science - The Reproducibility Project: Cancer Biology

This initiative was designed to provide evidence of reproducibility in cancer research and to identify possible factors that may affect reproducibility. Here, selected results from high-profile articles are independently replicated by unbiased third parties to evaluate if data could be consistently reproduced. For each evaluated study, a registered report delineating the experimental workflow is reviewed and published before experimentation is initiated; after data collection and analysis, the results are published as a replication study.

Author Policies for Publication

Many peer-reviewed journals have updated their reporting requirements to help improve the reproducibility of published results. The Nature Research journals, for example, have implemented new editorial policies that help ensure the availability of data, key research materials, computer codes and algorithms, and experimental protocols to other scientists. Researchers must now complete an editorial policy checklist to ensure compliance with these policies before their manuscript can be considered for review and publication.

Most people familiar with the issue of reproducibility agree that these efforts are gaining traction. However, progress will require sustained attention on the issue, as well as cooperation and involvement from stakeholders across various fields.

The academic research system encourages the rapid publication of novel results.

Moving forward

Accuracy and reproducibility are essential for fostering robust and credible research and for promoting

scientific advancement. There are predominant factors that have contributed to the lack of reproducibility in life science research. This issue has come to light in recent years and a number of guidelines and recommendations on achieving reproducibility in the life sciences have emerged, but the practical implementation of these practices may be challenging. It is essential that the scientific community are objective when designing experiments, take responsibility for depicting their results accurately, and thoroughly and precisely describe all methodologies used. Further, funders, publishers, and policy-makers should continue to raise awareness about the lack of reproducibility and use their position to promote better research practices throughout the life sciences. By taking action and seeking opportunities for improvement, researchers and key stakeholders can help improve research practices and the credibility of scientific data.

For more information on how you can improve the reproducibility of your research, visit ATCC online.

References

1. Ioannidis JP. *PLoS Medicine* **11**: e1001747, 2014.
2. Feilden T. Science & Environment, BBC News, February 22, 2017.
3. Baker M. *Nature News Feature*, May **25**, 2016.
4. ASCB. ASCB, 2014.
5. Freedman LP, Cockburn IM, Simcoe TS. *PLoS Biology* **13**: e1002165, 2015.
6. Chalmers I, Glasziou P. *Lancet* **374**: 86-89, 2009.
7. Macleod MR, *et al. Lancet* **383**: 101-104, 2014.
8. Horbach S, Halffman W. *PLoS One* **12**: e0186281, 2017.
9. Mouriaux F, *et al. Invest Ophthalmol Vis Sci* **57**(13): 5288-5301, 2016.
10. Liao H, *et al. Cytotechnology* **66**: 229-238, 2014.
11. Somerville GA, *et al. J Bacteriol* **184**: 1430-1437, 2002.

show less

12. Grimm D, *et al. Infect Immun* **71**(16): 3138-3145, 2003.

13. Lee JY, *et al. Scientific Reports* 6: 25543, 2016.

14. Resnik DB, Shamoo AE. *Account Res* **24**(2): 116-123, 2017.

15. The Academy of Medical Sciences, BBSRC, Medial Research Council, Wellcome Trust. Symposium report, October 2015.

16. Munafò MR, *et al. Nature Human Behaviour* 1: 0021, 2017.

17. Cherry K. Cognitive Psychology, Very Well Mind. October 8, 2018.

18. Stodden V, Leisch F, Peng RD. 1st Edition, Chapman and Hall/CRC, 2014.

19. Landis SC, *et al. Nature* **490**: 187-191, 2012.

20. NIH https://www.nih.gov/research-training/rigor-reproducibility 2015.

21. Davies, EW, Edwards, DD. A Report from the American Academy of Microbiology: Promoting Responsible Scientific Research, 2016.

22. Weintraub PG. *J Insect Sci* 16(1): 109, 2016.

Access to Research Data and EU Copyright Law

by **Linda Kuschel and Jasmin Dolling***

Abstract: With the advent of data-driven science and data-based business models in the 21st century, legal questions surrounding data, data rights and data law have become one of the most discussed topics both for lawmakers and for legal scholars globally. This is true particularly in the European Union, which in recent years has introduced data protection legislation, cybersecurity legislation, legislation regarding digital content and digital services, and more. Within this flurry of legal activity, one area of data law goes surprisingly unnoticed—the generation, ownership and use of research data. The slim attention it receives is disproportionate to its relevance in the digital economy. Not only are research data essential for the development of new technologies, they also feed machine-learning algorithms and are produced in any and all academic institutions. In order to maximize innovative potential, it is essential that researchers operate with legal certainty when using research data. The article seeks to contribute to this aim by exploring the legal framework in which research data can be accessed and used in EU copyright law. First, it delineates the authors' understanding of research data. It then examines the protection research data currently receives under EU and Member State law via copyright and related rights, as well as the ownership of these rights by different stakeholders in the scientific community. After clarifying relevant conflict-of-laws issues that surround research data, it maps ways to legally access and use them, including statutory exceptions, the open science movement and current developments in law and practice.

Keywords: Copyright; research data; freedom of science; open access; Open Science

Recommended citation: Linda Kuschel and Jasmin Dolling, Access to Research Data and EU Copyright Law, 13 (2022) JIPITEC 247 para 1.

A. Copyright and Research: Friends or Foes?

1 The rationale behind copyright law appears as relevant for research data as it is for creative works in the traditional sense: creativity should be incentivized while the embedded information should circulate and be disseminated as freely as possible. The temporary monopoly that copyright protection grants is not intended to deprive the general public of ideas, methods or doctrines,[1] as this would endanger the scientific communication process[2] and societal advancement. Why, then, does copyright often appear to get in the way of conducting research?

2 Arguably, copyright's focus has shifted from promoting intellectual creations towards protecting investments. The standard of creativity is low;[3] related rights grant protection to products such as audio recordings and photographs, which can contain no creativity of their own. The *sui generis* protection of databases, which stems from European

* Prof. Dr. Linda Kuschel, LL.M. (Harvard) holds a junior professorship in civil law, intellectual property law and law and digitization at Bucerius Law School, Hamburg. Jasmin Dolling is a graduate research assistant at Bucerius Law School, Hamburg. We thank Tarmio Frei for valuable research assistance.

1 BGHZ 39, 306 = FCJ 27 March 1963 – I b ZR 129/61 – NJW 1963, 1877, 1878 – *Rechenschieber*.

2 Cf Michael Fehling, 'Verfassungskonforme Ausgestaltung von DFG-Förderbedingungen zur Open-Access-Publikation' (2014) OdW 179, 189.

3 See further infra, B.I.

law, even rewards solely an investment effort. For researchers not versed in the terrain of copyright law, obligations when using or generating data have become increasingly unclear in the face of varied and ever-new types of protection, divergent requirements for protection between and sometimes even within jurisdictions and complex meshes of rightholders. The article therefore seeks to illuminate the role of research data and its useability in European copyright law,[4] presenting a definition of research data (B.), the types of protection they may enjoy (C.), common rightholders (D.), conflict-of-laws problems in international use (E.) and, finally, ways to legally access and use them, including statutory exceptions, the open science movement and current developments in law and practice (F.).

B. Research Data: An Attempt to Clarify

3 Before examining the legal questions that arise when using research data, one must delineate which types of data this term encompasses. A universal definition is not self-evident, as perspectives on what research data are, what form they take and which purpose they have differ between and sometimes even within scientific disciplines.[5] A natural and technical science approach, for example, might define research data as "experimental results, observations and computer-generated information[,] which form the basis for the quantitative analysis underpinning many scientific publications".[6] On the other end of the spectrum, there are initiatives like NFDI4Culture, a consortium for the digitization and integration of research data on material and immaterial cultural heritage. Their understanding of research data includes digital representations of cultural assets such as, eg, paintings, photographs and sketches, 3D models of buildings and musical or stage performances, as well as procedural research data resulting from research on these cultural assets, amongst others.[7]

4 In order to benefit different scientific disciplines, the term research data must therefore be interpreted broadly. For the purposes of this article, research data are thus understood to be objects of information subject to the scientific cognitive process.[8] They can exist at the outset of the research activity as well as be generated, or rather developed through interpretation during its course.

5 Further, the present consideration includes not only digital, but also analogue objects of information, such as handwritten notes or photographs. While it is indubitable that digitization gives rise to new possibilities of production, storage and analysis of research data, not all data are originally digital. Moreover, the assessment of research data's eligibility for copyright protection largely takes place irrespective of whether data exist in analogue or digital form.

C. Layers of Legal Protection

6 Research data can appear in many different shapes and sizes. The assessment of their legal protection (under copyright law) is invariably contingent on their specific manifestation.[9] In the following section, the ways in which research data and EU copyright protection intersect are presented by (I.) charting copyright's core, the protection of works, (II.) exploring related rights and (III.) outlining the *sui generis* right in databases, which indirectly includes even raw data.

4 Additionally, aspects of data (protection) law, patent law or trade secrets law can be of particular relevance. These are beyond the scope of the present article.

5 Cf Thomas Hartmann, 'Urheberrechtliche Schutzfähigkeit von Forschungsdaten' in Jürgen Taeger (ed), *Law as a Service. Recht im Internet- und Cloud-Zeitalter* (OlWIR 2013) 505, 508; Heinz Pampel, Hans-Jürgen Goebelbecker, Paul Vierkant, 're3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositorien. Ein Werkstattbericht' in Bernhard Mittermaier (ed), *Vernetztes Wissen. Daten, Menschen, Systeme* (Forschungszentrum Jülich 2012) 61, 62; Jakob Voß, 'Was sind eigentlich Daten?' (2013) 23 LIBREAS. Library Ideas 4, 6 <https://libreas.eu/ausgabe23/02voss/> accessed 14 March 2022.

6 European Commission, Towards better access to scientific information: Boosting the benefits of public investments in research, Communication from 17 July 2012, COM(2012) 401 final, 3.

7 Torsten Schrade, 'NFDI4Culture', (2022) 5 Politik & Kultur 7; Reinhard Altenhöner et al., 'NFDI4Culture - Consortium for research data on material and immaterial cultural heritage', (2020) 6 Research Ideas and Outcomes, <https://doi.org/10.3897/rio.6.e57036> accessed 5 July 2022.

8 This is in keeping with previous definitory endeavors, such as Art 2(9) Open Data Directive ("[D]ocuments in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results") or the definition posed by the Alliance of German Scientific Organizations' focus initiative digital information ("data generated in the course of scientific projects", cf <www.allianzinitiative.de/archiv/forschungsdaten/> accessed 8 June 2022).

9 BGHZ 112, 243 = FCJ 27 September 1990 – I ZR 244/88 – GRUR 1991, 523, 525 – *Grabungsmaterialien*.

I. Research data as protected works

7 Copyright protects the rights of authors for "works in the literary, scientific and artistic domain".[10] Thus, protection under copyright requires a work. While European copyright directives do not contain an express definition of the term "work", the Court of Justice of the European Union ("ECJ") has developed two conditions that must both be satisfied for copyright eligibility:[11] first, there must be an original subject matter, ie, the author's own intellectual creation; second, only the expression of this creation can be copyright-protected as a work.[12] Assessing whether and when these conditions are fulfilled has largely been left to the Member States to determine on a case-by-case basis; however, the ECJ has ruled on copyright protection in certain constellations, enabling general conclusions on the Court's understanding of these conditions. In *Brompton*, a case for copyright infringement of a folding bicycle able to take three distinct positions, the Court confirmed that copyright protection extends to products whose shape is at least partially necessary to obtain a certain technical result, so long as, through the shape, the author expresses their creative ability by making free and creative choices, so that the shape reflects their personality.[13] In *Cofemel*, respectively, the Court denied copyright protection for (in this case, clothing) designs that, beyond their practical purpose, generate only an aesthetically significant visual effect. The Court held that an aesthetic effect alone was not enough to determine whether a design constitutes an author's intellectual creation, and a subjective aesthetic effect further did not equate to an expression, ie, a subject matter that is existing and identifiable with sufficient precision and objectivity.[14] This case law allows two conclusions to be drawn for research data: first, while research data will often be of a technical, functional nature (eg, the results of a clinical trial or studies of a chemical reaction), this does not in principle exclude them from copyright protection; second, research data must reflect their author's creativity in order to be eligible for protection as copyrighted works.

8 In many academic disciplines, particularly the humanities, research is conducted by analyzing sources including literature, musical compositions, artistic works, photographic works or films, which will generally enjoy copyright protection if they are not already in the public domain. Copyright protection expires seventy years after the author's death (Article 1(1) Copyright Term Directive[15]), or after the death of the last surviving joint author (Article 1(2) Copyright Term Directive). Within this exclusive period, if research data consists of collected pre-existing material, such as literary texts or other creative content, it is likely to be copyright-protected and its use must be in accordance with statutory exceptions or contractual licenses.[16]

9 If research data does not consist of such material that clearly falls into the realm of copyright, its protection depends upon its form; research data that exist in the form of written text can be protected as literary works if they constitute the authors' own intellectual creations. While the threshold for the level of creativity in literary works is relatively low,[17]

10 Sec 1 *Urheberrechtsgesetz* 1965 (Copyright Act Germany); cf also Sec 1 no 2 *Zakon o autorskom pravu i srodnim pravima* 2003 (Copyright Act Croatia); Sec 2(1) *Autorský zákon* 2000 (Copyright Act Czech Republic); Sec 1(2) no 1 *Autoriõiguse seadus* 1992 (Copyright Act Estonia); Sec L112-2 *Code de la propriété intellectuelle* 1992 (Copyright Act France); Sec 2(1) Copyright Act Greece; Sec 1(1) *Törvény a szerzői jogról* 1999 (Copyright Act Hungary); Sec 1(1) *Autorių teisių ir gretutinių teisių* įstatymas 1999 (Copyright Act Lithuania); Sec 1 *Auteurswet* 1912 (Copyright Act Netherlands); Sec 1(1) no 1 *Zakon o avtorski in sorodnih pravicah* 1995 (Copyright Act Slovenia).

11 The ECJ's competency for establishing an autonomous, uniform definition of the term is debated amongst scholars (cf Mireille van Eechoud, 'Along the Road to Uniformity – Diverse Readings of the Court of Justice Judgments on Copyright Work' (2012) 3 JIPITEC 60, paras 90ff; Eva-Marie König, *Der Werkbegriff in Europa* (Mohr Siebeck 2015), 22ff; Haimo Schack, 'EuGH: Kein Urheberrechtsschutz *für* Lebensmittelgeschmack mangels Werkcharakter - Levola/Smilde' (2019) 1 GRUR 75 (note)). The Court deems itself competent because the relevant Directives, particularly Council Directive 2001/29/EC of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society [2001] OJ L167/10 (InfoSoc Directive), do not expressly place the subject matter into the scope of competency of the Member States (cf ECJ, Case C-310/17 *Hexenkaas* [2018] ECLI:EU:C:2018:899 para 33; ECJ, Case C-5/08 *Infopaq International* [2009] ECR I-06569 para 27).

12 Cf ECJ, Case C-683/17 *Cofemel* [2019] ECLI:EU:C:2019:721 para 29.

13 ECJ, Case C-833/18 *Brompton* [2020] ECLI:EU:C:2020:461 paras 23ff.

14 Cf ECJ, Case C-683/17 *Cofemel* [2019] ECLI:EU:C:2019:721 paras 53f.

15 Council Directive 2006/116/EC of 12 December 2006 on the term of protection of copyright and certain related rights [2006] OJ L372/12.

16 Note that if copyright in a previously unreleased work has expired and this work is then released or communicated to the public for the first time, an exclusive right of exploitation is granted for 25 years (Art 4 Copyright Term Directive).

17 FCJ 15 September 1999 – I ZR 57/97 – GRUR 2000, 144, 145

a text of certain length is generally required. At the same time, there is no fixed word or character limit,[18] theoretically even single sentences[19] or tweets[20] are eligible for protection if the author expresses themself in a particularly creative fashion. For research data in text form describing the results of an experiment, documenting an observation or annotating data for machine learning purposes, such a creative mode of expression will be unlikely and, in any case, hardly wanted.[21] However, creative efforts consisting in expressing complex facts as clearly and precisely as possible are also rewarded.[22] It follows that research data presented in a piece of writing may well enjoy copyright protection—although generally only where there is enough leeway to describe the results found in individual words, and only with regard to their creative expression.[23] Very brief texts predominantly composed of fixed terminology, such as anamnesis reports, are rather unlikely to merit protection. Moreover, the methods, theories and results expressed within the text remain copyright-free.[24]

10 Illustrations of a scientific or technical nature can also be protected works.[25] The examples provided in many of the statutes[26] (in Germany, eg, "drawings, plans, maps, sketches, tables and three-dimensional representations") are visualizations often used in connection with research data. As with texts, a sufficient measure of creative expression is required. This measure is not reached where the representation is purely schematic and dictated by scientific norms.[27] However, copyright protection is not precluded by the content of a representation being of technical nature and the information being presented as clearly as possible.[28]

11 Collections and databases also enjoy copyright protection. Protected collections are "[c]ollections of literary or artistic works such as encyclopaedias and anthologies which, by reason of the selection or arrangement of their contents, constitute intellectual creations [...]."[29] Thus, protection arises from the particular selection and arrangement of elements, not from their content, and relates only to that specific selection and arrangement.[30] The same applies to database works eligible for protection under Article 3(1) Database Directive.[31] They are collections

– *ComicÜbersetzungen II*.

18 Cf Winfried Bullinger, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), § 2 paras 27f; Axel Nordemann, *Urheberrecht* (Axel Nordemann, Jan Bernd Nordemann et al. eds, 12th edn, W. Kohlhammer 2018), § 2 para 59; Haimo Schack, *Urheber- und Urhebervertragsrecht* (9th edn, Mohr Siebeck 2019), para 202; Ulrich Loewenheim and Matthias Leistner, *Urheberrecht* (Ulrich Loewenheim, Matthias Leistner and Ansgar Ohly eds, 6th edn, C.H. Beck 2020), § 2 para 45.

19 Regional Court of Munich 8 September 2011 – 7 O 8226/11 – GRUR-RR 2011, 447 – *Karl Valentin*. Cf also ECJ Case C-5/08 *Infopaq* [2009] ECR I-06569 paras 47f.

20 Higher Regional Court of Cologne 8 April 2016 – 6 U 120/15 - K&R 2016, 423; Regional Court of Bielefeld 3 January 2017 – 4 O 144/16 - MMR 2017, 641 (in these specific cases, protection was denied for lack of the required level of creativity). Cf also Gernot Schulze, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 2 para 83; Hannes Ludyga, 'Urheberrechtlicher Schutz von Tweets' (2017) 48 AfP 284.

21 Cf Hartmann (n 5) 511; Nordemann (n 18) para 118; Schulze (n 20) para 93. Critical Helmut Haberstumpf, 'Wem gehören Forschungsergebnisse?' (2001) 11 ZUM 819, 821.

22 FCJ 11 April 2002 – I ZR 231/99 – GRUR 2002, 958, 959 – *Technische Lieferbedingungen*.

23 FCJ 21 November 1980 – I ZR 106/78 – GRUR 1981, 352, 353 – *Staatsexamensarbeit*. Cf also Bullinger (n 18) para 57; Schulze (n 20) para 93.

24 BGHZ 39, 306 = FCJ 27 March 1963 – I b ZR 129/61 – NJW 1963, 1877, 1878 – *Rechenschieber*; FCJ 21 November 1980 – I ZR 106/78 – GRUR 1981, 352, 353 – *Staatsexamensarbeit*. Cf also Bullinger (n 18) para 50; Loewenheim (n 18) para 71; Schulze (n 20) para 93.

25 Cf eg Sec 2 no 3 *Urheberrechtsgesetz* 1936 (Copyright Act Austria); Sec 4(3) no 2 Copyright Act Estonia; Sec L112-2 no 12 Copyright Act France; Sec 2(1) no 7 Copyright Act Germany; Sec 2(1) Copyright Act Greece; Sec 5(2) no 12 Copyright Act Slovenia.

26 Sec 2(1) no 7 Copyright Act Germany. Cf also Sec 5(2) Copyright Act Croatia; Sec 4(3) no 2 Copyright Act Estonia; Sec L112-2 no 12 Copyright Act France; Sec 2(1) Copyright Act Greece; Sec 4(2) no 21 Copyright Act Lithuania; Sec 5(2) no 12 Copyright Act Slovenia.

27 Hartmann (n 5) 511.

28 FCJ 10 May 1984 – I ZR 85/82 – GRUR 1985, 129, 130 – *Elektrodenfabrik*.

29 Art 2(5) Berne Convention for the Protection of Literary and Artistic Works. Cf also Sec 6 Copyright Act Austria; Sec 7(1) Copyright Act Croatia; Sec 4(3) no 22 Copyright Act Estonia; Sec L112-3 Copyright Act France; Sec 4(1) Copyright Act Germany; Sec 2(2) Copyright Act Greece; Sec 7(1) Copyright Act Hungary; Sec 4(3) no 3 Copyright Act Lithuania; Sec 10(2) Copyright Act Netherlands.

30 Cf FCJ 7 December 1979 – I ZR 157/77 – GRUR 1980, 227, 230f – *Monumenta Germaniae Historica*.

31 Council Directive 96/9/EC of 11 March 1996 on the legal

"of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means" (Article 1(2) Database Directive). In the context of research, a collection of raw data, such as measurement data from a test series, will generally be comprehensive in nature and therefore not subject to an individual selection; the arrangement in turn will follow logical criteria (eg, time, quantity, size), as the representation should meet scientific standards and be as clear and verifiable as possible. Therefore, there is little room for creative selection or arrangement.[32] The case may be different, eg, in the humanities or cultural studies, where a research database could consist of a selection of poetry[33] (based on individual criteria[34]) or newspaper articles.[35] Yet, the investment or work effort put into a database or the expertise necessary cannot be taken into account in the question of whether a research database constitutes an intellectual creation.[36] However, they play a role in the related *sui generis* right in databases.[37]

12 Computer programs can also be protected by copyright, provided they contain the programmers' own intellectual creation and, as such, reflect a minimum of individuality.[38] Entirely trivial program designs or pre-existing program elements, therefore, are not protected.[39] In any case, protection arises only for the expression of the program, not for its underlying ideas and principles (cf Article 1(2) Computer Programs Directive).[40]

II. Related rights to research data

13 While copyright is granted only for intellectual creations, related rights extend to certain non-creative efforts related to copyright-protected works. For research data, the related rights to photographs and moving pictures and the protection of producers of audio recordings play a particular role.[41]

14 In a number of Member States, photographs and "products manufactured in a similar manner to photographs" are protected.[42] The term photograph encompasses any type of photography, irrespective of its specific imaging technology, and therefore includes, for example, aerial and satellite photographs.[43] Products manufactured in a similar manner are all images produced using radiant energy.[44] These include, eg, infrared images, medical x-ray or ultrasound images, as well as magnetic resonance or computer tomography images, which are particularly relevant for research data.[45]

15 Moreover, sequences of images or sequences of images and sounds that are not protected as cinematographic works, ie, which do not fulfil the requirements for copyright protection, can still receive protection as moving pictures in two Member States.[46] Typically, these are films that merely docu-

protection of databases [1996] OJ L77/20.

32 This view is also supported by Fehling (n 2) 188; Hartmann (n 5) 512; Gerald Spindler, 'KoLaWiss-Gutachten AP 4: Recht' (2009), 30ff.

33 BGHZ 172, 268 = FCJ 24 May 2007 – I ZR 130/04 – NJW 2008, 755, 756 – *Gedichttitelliste I*.

34 Cf Sören Rieger, *Der rechtliche Schutz wissenschaftlicher Datenbanken* (Mohr Siebeck 2010) 101.

35 Higher Regional Court of Hamm 26 February 2008 – 4 U 157/07 – ZUM 2008, 598, 601.

36 ECJ, Case C-604/10 *Football Dataco et al.* [2012] ECLI:EU:C:2012:115 para 42. Cf also Eva-Marie König, *Der Werkbegriff in Europa* (Mohr Siebeck 2015), 18f; Thomas Dreier, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 4 para 11.

37 See infra, C.III..

38 Cf Eva-Marie König, *Der Werkbegriff in Europa* (Mohr Siebeck 2015), 16f; Nordemann (n 18) para 75.

39 FCJ 20 September 2012 – I ZR 90/09 – GRUR 2013, 509 para 25 – *UniBasic–JDOS*; Higher Regional Court of Berlin 6 September 2010 – 24 U 71/10 – ZUM-RD 2011, 544, 547 – *FRITZ!Box*. Cf also Schulze (n 20) para 127.

40 Council Directive 2009/24/EC of 23 April 2009 on the legal protection of computer programs [2009] OJ L111/16.

41 See infra, C.III., on the *sui generis* protection of makers of database.

42 This protection is not mandated by EU law, cf Art 6 third sentence Copyright Term Directive. It is granted in Austria (Sec 73 Copyright Act Austria), Denmark (Sec 70 *Ophavsretsloven* 2014 (Copyright Act Denmark)), Germany (Sec 72 Copyright Act Germany), Finland (Sec 49a *Tekijänoikeuslaki* 1961 (Copyright Act Finland)), Spain (Sec 128 *Ley de Propiedad Intelectual* 1996 (Copyright Act Spain)) and Sweden (Sec 49a *Upphovsrättslagen* 1960 (Copyright Act Sweden)).

43 Cf Gernot Schulze, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 72 para 3f.

44 Cf Schulze (n 43) para 6; Dorothee Thum, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), § 72 para 24.

45 Cf Schulze (n 43) para 6; Thum (n 44) para 24.

46 In Austria, Secs 73(2), 74 Copyright Act Austria, and Germany,

ment an event or a process without employing tools of creative cinematic design.[47] Research data that, for example, capture a test procedure, a natural event or an interview on film, are therefore protected as moving pictures (alongside the protection of the individual film frames as photographs).

16 The most important difference between the protection of photographs or films as works and the related rights for photographs or moving pictures is that the latter do not depend on creative expression and extend to faithful, objective reproductions of events.[48] Therefore, research data in form of images and films can usually elicit protection as photographs or moving pictures (only). The scope of related rights for photographs and moving pictures differs from the protection of copyrighted works primarily through a shorter term of protection granted by the Member States—50 years after publication (or production) rather than 70 years *post mortem auctoris* (pma).[49] In addition, there is a particularity for photographs of works of visual art that are in the public domain. For a long time, it was controversial whether photographic replications of two-dimensional originals, especially photographs faithful to an original painting, could enjoy protection under copyright law.[50] Notably, the legal setting has changed after the adoption of the DSM Directive[51], effective June 6, 2019 (the transposition period ended June 7, 2021). The Directive establishes in its Article 14 that "when the term of protection of a work of visual art has expired, any material resulting from an act of reproduction of that work is not subject to copyright or related rights, unless the material resulting from that act of reproduction is original in the sense that it is the author's own intellectual creation". This precludes photographic replications of works of visual art in the public domain from being protected as photographs; they may only enjoy copyright protection as photographic works if the associated higher standards of creative expression are fulfilled (see supra, B.I.). For research data, this will predominantly not be the case.

17 The Member States can provide for a related right to critical and scientific publications, which is granted for publications that contain works no longer protected by copyright.[52] In Germany, this right is also granted for scientific publications of unprotected texts, such as letters, maps or judicial proceedings, and for works not protected by copyright for other reasons.[53] It requires, however, that there has been scientifically organized activity and (in case the works or texts contained therein, which have been published previously) that they differ significantly from previous editions of the works or texts.[54] Scientifically organized activity requires sighting, organizing and evaluating work, employing scientific methods.[55] The right can be granted for 30 years at most.

18 Finally, research data can exist in the form of audio recordings, such as of interviews, group discussions or nature sounds. Regardless of a possible copyright protection (for example in the case of a creative speech that has been recorded), the production of the audio recording as such is protected under Article 2 lit c InfoSoc Directive.[56] As opposed to the speaker's

Sec 95 Copyright Act Germany. Cf on this Günter Poll, 'Die Harmonisierung des europäischen Filmurheberrechts aus deutscher Sicht' (2003) 4 GRUR Int 290, 293f.

47 Cf Schack (n 18) para 730.

48 Cf Thum (n 44) para 22.

49 Cf for photographs, Sec 74(6) Copyright Act Austria, Sec 70(2) Copyright Act Denmark, Sec 72(3) Copyright Act Germany, Sec 49a(2) Copyright Act Finland, Sec 49a(2) Copyright Act Sweden; for moving pictures, Sec 95 in conjunction with 94(3) Copyright Act Germany, Secs 73(2), 74(6) Copyright Act Austria. This is in synchronicity with the protection terms for related rights established in Art 3 Copyright Term Directive.

50 Cf Schulze (n 43) para 10; Thum (n 44) para 23. This is because even protection only as a photograph must demonstrate at least a small minimum of own intellectual (but not creative) effort, which is not present, eg, in reproductive photocopies or scans (Schack (n 18) para 722; Schulze (n 43) para 9). This fundamental concept was confirmed by the German Federal Court of Justice in its decision *Museumsfotos* (FCJ 20 December 2018 – I ZR 104/17 – GRUR 2019, 284, 286 para 23). In the case at hand, however, the FCJ considered the photographers' decisions on lighting, angle and distance to the photographed painting to be a sufficient intellectual effort for protection as a photograph. According to these standards, research data depicting two-dimensional works (in the public domain), such as in art or media studies, would generally enjoy protection as photographs.

51 Council Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market [2019] OJ L130/92.

52 Art 5 Copyright Term Directive. This right has been introduced in Estonia and Germany, cf Sec 74[1](2) Copyright Act Estonia; Sec 70 Copyright Act Germany.

53 Anne Lauber-Rönsberg, *Urheberrecht* (Hartwig Ahlberg, Horst-Peter Götting and Anne Lauber-Rönsberg eds, 33rd edn, C.H. Beck 2022), § 70 para 5.

54 Cf Sec 70(1) Copyright Act Germany.

55 FCJ 23 May 1975 – I ZR 22/74 – GRUR 1975, 667, 668 – *Reichswehrprozess*.

56 Council Directive 2001/29/EC of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society [2001] OJ L167/10.

copyright, the rights of the audio recording's producer expire 50 years after the fixation or, in case of lawful publication or communication to the public within this period, 50 years from the date of the first act of publication or communication, Article 3(2) Copyright Term Directive.

III. Protection of research data in databases

19 The previous sections dealt with research data that have, in different ways, taken a creative form. Raw data, meaning non-edited data such as measurements, do not enjoy copyright protection as such. They can, however, constitute a protected database under Article 3(1), 1(2) Database Directive insofar as they exist in larger number and are ordered systematically.

20 The *sui generis* right for the maker of a database encompasses any "collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means" (Article 1(2) Database Directive) and the "obtaining, verification or presentation" of which requires "qualitatively and/or quantitatively a substantial investment" (Article 7(1) Database Directive). Usually, a research database is a collection of works, data or other individually (electronically or otherwise) accessible elements. "Independency" requires that the elements can be separated from each other without adversely affecting the value of their content.[57] This is intended to prevent an extension of the term "database" to include all items composed of individual components (such as musical compositions, which are made up of musical notes). The elements must make sense independently, not only in their combination.[58] However, the ECJ applies a rather generous standard: the individual data arising from a topographic map (terrain altitude, location of traffic roads etc.) are sufficiently independent, even if the purpose of a map unfolds only in viewing all its elements in combination.[59] For research data, this means that not only a collection of different data, but a single document (such as a drawing of an archaeological excavation site) may already constitute a database—provided that it fulfils the other requirements for protection.

21 The requirement of a systematic or methodical arrangement is intended to distinguish a database from a mere collection of raw data not compiled by organizational criteria.[60] Since research data are compiled according to plausible organizational criteria in order to ensure their scientific useability, this prerequisite is easily fulfilled.

22 Finally, the database must show that there has been a substantial qualitative or quantitative investment. The *sui generis* database right was intended to reward the investment effort of a database producer, thereby creating an incentive to develop "modern information storage and processing systems".[61] An investment cannot only be the expenditure of money, but also of time, work or technical means.[62] In four judgments from November 9, 2004, the ECJ clarified that only investments associated with the creation of the database, ie, obtaining, verifying or presenting the data, are relevant.[63] In this respect, investments serving the generation of data are not considered.[64] For research data, a—not entirely

57 ECJ, Case C-604/10 *Football Dataco et al.* [2012] ECLI:EU:C:2012:115 paras 26f; ECJ, Case C-203/02 *The British Horseracing Board et al.* [2004] ECR I-10415 para 31. Cf also Kirsten Johanna Schmidt and Herbert Zech, 'Datenbankherstellerschutz für Rohdaten?' (2017) 33 CR 417, 418.

58 Schmidt and Zech (n 57) 419.

59 ECJ, Case C-490/14 *Freistaat Bayern v Verlag Esterbauer GmbH* [2015] ECLI:EU:C:2015:735 paras 25f. Critical Matthias Leistner, 'Was lange währt…: EuGH entscheidet zur Schutzfähigkeit geografischer Karten als Datenbanken' (2016) 1 GRUR 42.

60 Cf Rec 21 Database Directive; ECJ, Case C-444/02 *Fixtures Marketing II* [2004] ECR I-10549 para 30; Schmidt and Zech (n 57) 420; Martin Vogel, *Urheberrecht* (Ulrich Loewenheim, Matthias Leistner and Ansgar Ohly eds, 6th edn, C.H. Beck 2020), § 87a para 22.

61 Cf Rec 12 Database Directive.

62 Cf FCJ 1 December 2010 – I ZR 196/08 – GRUR 2011, 724, 725, para 18 – *Zweite Zahnarztmeinung II*; Estelle Derclaye, 'Database Sui Generis Right: What Is a Substantial Investment? A Tentative Definition' (2005) IIC 4ff; Thomas Dreier, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 87a para 12; Schmidt and Zech (n 57) 421.

63 ECJ, Case C-203/02 *The British Horseracing Board et al.* [2004] ECR I-10415, para 31; ECJ, Case C-338/02 *Fixtures Marketing I* [2004] ECR I-10497 para 24; ECJ, Case C-444/02 *Fixtures Marketing II* [2004] ECR I-10549 para 40; ECJ, Case C-46/02 *Fixtures Marketing III* [2004] ECR I-10365 paras 31ff.

64 ECJ, Case C-203/02 *The British Horseracing Board et al.* [2004] ECR I-10415 para 31; ECJ, Case C-338/02 *Fixtures Marketing I* [2004] ECR I-10497 para 24; ECJ, Case C-444/02 *Fixtures Marketing II* [2004] ECR I-10549 para 40; ECJ, Case C-46/02 *Fixtures Marketing III* [2004] ECR I-10365 paras 31ff. Cf also Matthias Leistner, 'ECJ, Case C-203/02 *The British Horseracing Board et al.* [2004] ECR I-10415' (2005) JZ 408, 409 (note).

trivial—distinction between investments in data generation and investments in data obtainment and collection must therefore be made.[65]

23 Partially, the distinction is made by separating data that are "found", ie, pre-existing data and data that are "invented".[66] Only the latter are said to carry the danger of monopolizing information, as the inventor of the data is in the exclusive position to collect them.[67] On the other hand, data that are collected as part of scientific measurements or observations are said to be pre-existing in nature and the expenditure to collect them therefore to constitute an eligible investment.[68] The argument made against this criterion is that in nature, only "potentially semantic information" pre-exists, which needs human perception to be turned into de facto information and, thereby, into data.[69] Indeed, it does not seem entirely plausible to classify factual events as "pre-existing data" before they are documented. Instead, it is proposed to use a criterion of general perpetual accessibility, ie, to ask whether "third parties could, with similar expenditure, create the same data".[70] In case of ephemeral events, such as weather data, the observations could not be replicated and are therefore not perpetually accessible.[71] In principle, these criteria are persuasive and consistent, but they are not free of concerns; at least in the natural sciences, it is questionable whether a parallel observer would, in fact, identify "the same data". Surely, they would concede similar or equal results, but it is uncertain whether the data would be exactly identical. Moreover, the German Federal Court of Justice (FCJ) decided in the context of the motorway tolling system that the data collected within its framework, such as date and duration of drives subject to tolling, are "not created, but only collected and arranged".[72] However, these traffic data are also dynamic processes that can be assessed only in one specific moment. Therefore, the characterization of data as "found" or "invented" is highly difficult. Yet, it is important to keep in mind that, for the database protection right, it is not the nature of the data that is decisive, but rather which kind of investment was made. All investments that are necessary to initiate collectible information, ie, to launch a procedure that leads to information, are not relevant. For measurement data obtained from an experiment in the natural sciences, we must therefore differentiate between investments in the experimental setup and investments in measuring the experimental procedure and result. Only the latter can be taken into account for the *sui generis* database right. The same distinction can be made in social science experiments: costs for mobilizing experimental subjects (eg, recruitment) and the organizational planning of the experiment are irrelevant, while the expenditure of time by researchers documenting the procedures is to be included. Expenses for the processing of collected raw data constitute another area of investment; they are relevant costs for presentation of the data.[73]

24 Not to be included—at least under the aforementioned ECJ case law[74]—are investments in generating "synthetic data". Synthetic data are created in the context of machine learning algorithms with the aim to counterbalance misrepresentations in a given dataset. Over- or under-representation of individual groups (eg, in terms of gender or ethnicity) in a dataset can lead to discriminatory decisions or findings by the algorithm and thus has to be balanced out.[75] Since this kind of adjustment in training data is important as well as desirable, one might consider

65 Cf on this Rieger (n 34) 142ff.

66 Kai Hermes, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), § 87a paras 49ff; Vogel (n 60) para 53; Leistner, '*The British Horseracing Board et al.*' (n 64) 409; Matthias Leistner, 'Datenbankschutz. Abgrenzung zwischen Datensammlung und Datengenerierung' (2018) 34 CR 17, 20.

67 Leistner, 'Datenbankschutz' (n 66) 20.

68 Leistner, '*The British Horseracing Board et al.*' (n 64) 409.

69 Schmidt and Zech (n 57) 422. On this problem, cf also Timo Ehmann, *Wettbewerbsfreiheit und Investitionsschutz für Datenbanken* (C.H. Beck 2011), 109f.

70 Schmidt and Zech (n 57) 422.

71 Schmidt and Zech (n 57) 422.

72 FCJ 25 March 2010 – I ZR 47/08 – GRUR 2010, 1004, 1005 para 19 – *Autobahnmaut*.

73 Leistner, 'Datenbankschutz' (n 66) 19f. Appearing to dissent, Higher Regional Court of Hamburg 8 June 2017 – 5 U 54/12 – BeckRS 2017, 138204 para 247.

74 Philipp Hacker, 'Immaterialgüterrechtlicher Schutz von KI-Trainingsdaten' (2020) 10 GRUR 1025, 1030 argues for an interpretation of Art 7(1) Database Directive in light of Art 20 of the Charter of Fundamental Rights of the European Union [2012] OJ C326/391, that would allow for the inclusion of investments made in synthetic data for the purposes of balancing a given dataset.

75 This is also listed as one potential requirement related to training data by the EU Commission in its "White Paper on Artificial Intelligence" (COM(2020) 65 final, 19).

incentivizing respective investments by way of the database *sui generis* right, which is currently under revision.[76]

25 As opposed to database *works*, the selection and arrangement of individual elements requires no intellectual, creative effort to award the protection to database producers. It follows that, for research data, the *sui generis* right is much more relevant.[77] The duration of protection, however, is shorter: the protection for databases expires 15 years after publication or 15 years after production of the database, if it was not published within that period (Article 10(1), (2) Database Directive).

26 Similar to collections of works and database works, the protection for databases relates not to the individual data, but rather to the overall result, ie, the database. Therefore, the rightholder has the exclusive right to reproduce, distribute or make publicly available the database as a whole or "a qualitatively or quantitatively substantial part" of it (Article 7(1) Database Directive). However, the use of an insubstantial part of the database can already infringe the database producer's right if it is a "repeated and systematic" act that "runs contrary to a normal utilization of the database or unreasonably impairs the legitimate interests of the producer of the database" (Article 7(5) Database Directive). This "circumvention clause"[78] is intended to prevent systematic access to insubstantial parts of the database resulting in a prohibited use of a substantial part of or even of the entire database.[79] In its 2021 *Melons* judgment, the ECJ clarified that the use of a protected database is infringing if it adversely affects the database maker's investment in obtaining, verifying or presenting the content of the database (ie, constitutes a risk to the possibility of redeeming the investment through the normal operation of the database).[80] This follows from balancing the interests of the parties involved, in order to foster innovation and avoid a too far-reaching right of exclusivity for database makers.

27 The database right shall be further clarified by the proposed Data Act,[81] in which Article 35 holds that the sui generis right "does not apply to databases containing data obtained from or generated by the use of a product or related service" (see below, F.III.).

D. Rights to Research Data: Who and how many?

28 The previous section has shown that research data is almost always protected in some way; due to the rather extensive term of protection, only a fraction of research material is in the public domain. Access to and use of research data is further complicated by uncertainties about ownership and exclusive rights.

I. Authorship and original ownership

29 Particularly in the legal history of continental European copyright systems, the work's author takes center stage.[82] At least initially, copyright law has the genius (single) creator in mind,[83] whom it awards the exclusive rights to their work.[84] If multiple people have created the work, they also hold copyright jointly.[85] The same applies to related rights; here,

76 European Commission, *Communication, Making the most of the EU's innovative potential – An intellectual property action plan to support the EU's recovery and resilience* (25 November 2020), COM(2020) 760 final, 14f; European Commission, Communication, *Commission Work Programme 2021* (19 October 2020), COM(2020) 690 final, Annex I, 6b.

77 Cf Hartmann (n 5) 513.

78 ECJ, Case C-203/02 *The British Horseracing Board et al.* [2004] ECR I-10415 paras 84ff.

79 ECJ, Case C-203/02 *The British Horseracing Board et al.* [2004] ECR I-10415 paras 84ff. Cf also Kai Hermes, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), § 87b para 66.

80 ECJ, Case C-762/19 *CV-Online v Melons* [2021] ECLI:EU:C:2021:434 para 47.

81 European Commission, Data Act Proposal, 23 February 2022, COM(2022) 68 final.

82 Cf only Walter Bappert, *Wege zum Urheberrecht* (V. Klostermann 1962) 105ff; Martha Woodmansee, 'The Genius and the Copyright. Economic and Legal Conditions of the Emergence of the "Author"' (1984) 17 Eighteenth-Century Studies 424; Swedish Royal Commission, *Report on the Copyright to Literary and Artistic Works Bills*, SOU 1956:25 p 85 (authors' works are "their spiritual child").

83 This modern, auctorocentric notion of authorship arose in the eighteenth century in response to authors seeking to proprietarize their then expanding livelihood; this presents a break from the previous view of the author as an instrument, either of the court by which they were employed or of divine powers (Woodmansee (n 82) 425).

84 Cf for example in Austria, Sec 10(1) Copyright Act Austria, in Denmark, Sec 1(1) Copyright Act Denmark, in Estonia, Sec 28(1), (2) Copyright Act Estonia, in Finland, Sec 1(1) Copyright Act Finland, in Germany, Sec 7 Copyright Act Germany, in Ireland, Sec 21 *Copyright and Related Rights Act* 2000 (Copyright Act Ireland).

85 Cf eg Sec 11(1) Copyright Act Austria, Sec 6 Copyright Act

the person that has made the effort is awarded the right.[86] For films, the person that has made the economic and organizational effort of producing the film is entitled to the rights thereto (Article 2 lit d InfoSoc Directive)—the same applies to moving pictures by virtue of Member State legislation.[87] For research data generated by multiple people, the creator or producer of every element must be assessed individually.[88]

30 As seen, the *sui generis* database right is based on the hypothesis that the promise of legal protection furthers investments in the arrangement and structuring of data. Accordingly, the database producer, ie, the person that has made the substantial investment, is awarded the rights to a database eligible for protection, Article 7(1) Database Directive.[89] For university research, this is generally the university itself or a third party in case of funded or commissioned research.[90] Yet, the group of persons eligible for the *sui generis* database protection is restricted in an important way: only nationals of an EU Member State or persons whose habitual residence is located therein as well as companies and firms formed in accordance with the law of a Member State and having their registered office, central administration or principal place of business within the Community benefit from the database right (Article 11(1) and (2) Database Directive).

31 The basic assumption of copyright law, being that individual people create works, often fails to reflect the reality of large research projects.[91] This is because here, *groups of researchers* generally manage the project, and many different people participate in the generation of research data. Besides one or multiple group leaders, doctoral candidates and perhaps also research assistants and non-academic personnel often participate in a project. When the materials protected under copyright or related rights are assembled in a large database, a conglomerate is created, to which many different people have rights.[92]

II. Derivative rights

32 Legal systems in the Anglo-American copyright tradition allow not only for the transfer of copyright[93] but also for the initial ownership of an employer (work made for hire-doctrine).[94] Where these copyright regimes acknowledge moral rights,[95] they are—although not alienable—waivable.[96] Due to their roots in the right of personality, continental European author's rights systems, on the other hand, take, a different approach. Copyright consists of economic and moral rights that are always vested

Denmark, Sec 30(1) Copyright Act Estonia, Sec 6 Copyright Act Finland, Sec 8(1), (2) Copyright Act Germany, Sec 7(1) Copyright Act Greece, Sec 22(1), (4) Copyright Act Ireland, Sec 12(1) Copyright Act Slovenia.

86 Cf Thomas Dreier, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), Pre §§ 70ff para 13; Justine Pila and Paul LC Torremans, *European Intellectual Property Law* (1st edn, Oxford University Press 2016), 294.

87 Cf Sec 95 in conjunction with Sec 94 Copyright Act Germany; Sec 73(2) in conjunction with Sec 74(1) Copyright Act Austria. Cf also FCJ 6 February 2014 – I ZR 86/12 – GRUR 2014, 363, 364, para 23; Gernot Schulze, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 95 para 2.

88 Cf BGHZ 112, 243 = FCJ 27 September 1990 – I ZR 244/88 – GRUR 1991, 523, 525 – *Grabungsmaterialien;* Bernhard Ulrici, 'Kooperation in der Wissenschaft: Das Recht am und auf das Arbeitsergebnis' (2015) 48 WissR 318, 319f; Bernhard Ulrici, 'Geistiges Eigentum in Forschungsverbünden' (2018) OdW 129, 131.

89 Matthias Leistner, 'Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform' (7 September 2018), 5ff, <https://ssrn.com/abstract=3245937> accessed 19 September 2022; Michael Beurskens, 'Schranken des rechtlichen Schutzes von Datenbanken (Balancing Public and Private Interests in Database-Protection)', *Center for Business & Corporate Law Research Paper Series No. 0003* (21 November 2004), 51ff, <https://ssrn.com/abstract=646664> accessed 19 September 2022.

90 Anne Lauber-Rönsberg, Philipp Krahn and Paul Baumann, 'Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements' (2018), 5.

91 Cf on this Florian Möslein, 'Privatrechtliche Regelsetzungsfragen der wissenschaftlichen Kooperationsform: Angebot des Gesetzgebers oder selbstgestaltetes Recht?' (2018) OdW 99, 101.

92 Cf on the organization of access to research and research data generated during academic research projects *infra*, F.II.

93 Sec 120(1) Copyright Act Ireland; Sec 90(1) *UK Copyright, Designs and Patents Act* 1988 (Copyright Act UK).

94 Sec 23(1)(a) Copyright Act Ireland; Sec 11(2) Copyright Act UK. Notably, this rule is also employed in Sec 7 Copyright Act Netherlands.

95 Cf Calvin D Peeler, 'From the Providence of Kings to Copyrighted Things (and French Moral Rights)' (1999) 9(2) Ind Int'l & Comp L Rev 423; Cyrill P Rigamonti, 'Deconstructing Moral Rights' (2006) 47(2) Harvard Int L J 353; Stig Strömholm, 'Droit Moral – The International and Comparative Scene from a Scandinavian Viewpoint' (2002) 42 Scandinavian Stud L 217.

96 Sec 116(1) Copyright Act Ireland; Sec 87(2) Copyright Act UK.

in the author and may not be waived.[97] However, a closer examination reveals that the author's rights systems differ, too. In Germany, for example, copyright is construed monistically, economic and moral rights being an integrated whole, and the transfer of copyright[98] is thus precluded per se.[99] In France, commercial rights and moral rights are seen to be two separate pillars of copyright law, only the latter being inextricably linked to the creator's person, the former transferable.[100] However, even in countries where copyright is not transferable, third parties can be granted a right of use to the work, which may be extensive. Rights can either be granted explicitly via a contract of rights of use or arise implicitly from a private-law employment relationship or a public-law service relationship.[101] A number of particularities arise for (copyright-protected) research data produced at academic institutions.

1. Academic professors

33 The freedom of science guaranteed in many Member States' constitutions declares scientific research to be an autonomous area, free of government control,[102] so as not to endanger the role of research and teaching in furthering progress and understanding.[103] From this, we can deduce that the general rule for works created during the course of employment in most Member States, according to which the employer obtains rights of use in works created by the employee or is considered as their owner,[104] cannot be applied to works of academic professors without restrictions.[105] Other independently working researchers, such as private lecturers, adjunct professors or visiting lecturers, must be equated to academic professors.[106]

34 In this regard, Latvia has taken a pioneering role by introducing a Law on Scientific Activity stipulating that scientists, including academic professors, hold the exclusive rights to their research, insofar as there is no contractual agreement to the contrary.[107] While other states have similar statutes applicable to patents or utility models,[108] an explicit regulation for other IP rights to scientific research has thus far been lacking.

97 Cf Schack (n 18) paras 343f, 1114f.

98 In the majority view, the same applies to photographs. For German law cf Schulze (n 43) para 16; Thum (n 44) para 125. The rights of the producers of audio recordings (Sec 85(2) first sentence Copyright Act Germany) and of producers of films (Sec 95 in conjunction with 94(2) first sentence Copyright Act Germany), on the other hand, are transferrable.

99 Sec 29(1) Copyright Act Germany. Cf also Sec 23(3) Copyright Act Austria, Sec 42(1) Copyright Act Croatia, Sec 9(1), (3) Copyright Act Hungary.

100 Secs L121-1, L122-7 Copyright Act France. Cf also Sec 11(2) and (3) Copyright Act Estonia in conjunction with Sec 39 Constitution of the Republic of Estonia, Sec 12 Copyright Act Greece, Secs 14, 38 Copyright Act Lithuania; Sec 70 Copyright Act Slovenia.

101 Sec L113-9 Copyright Act France; Sec 43 Copyright Act Germany and, for computer programs, Sec 69b Copyright Act Germany; Sec 9(2) Copyright Act Lithuania.

102 Cf Sec 68 *Ustav Republike Hrvatske* 1990 (Croatian Constitution); Sec 77 *Danmarks Riges Grundlov* 1953 (Danish Constitution); Sec 38 *Eesti Vabariigi põhiseadus* 1992 (Estonian Constitution); Sec 5(3) *Grundgesetz* 1949 (German Constitution); Sec X *Magyarország alaptörvénye* 2011 (Hungarian Constitution); Sec 113 *Satversme* 1922 (Latvian Constitution); Sec 42 *Lietuvos Respublikos Konstitucija* 1992 (Lithuanian Constitution); Ch 2 Sec 18 *Regeringsformen* 1974 (Swedish Instrument of Government).

103 For Germany, this follows from BVerfGE 35, 79, 113 = German Federal Constitutional Court 29 May 1973 – 1 BvR 424/71 and 325/72 – NJW 1973, 1176. Cf on this Klaus F Gärditz, 'Die grundrechtliche Stellung der Wissenschaftlerinnen und Wissenschaftler in der Hochschulorganisation' (2016) 49 WissR 349, 357f.

104 Jurisdictions that transfer rights of use include Lithuania, Slovenia, Germany and Hungary. The rights of use may be obtained for a limited time (eg for 5 years in Lithuania, Sec 9 Copyright Act Lithuania, or for 10 years in Slovenia, Sec 101 Copyright Act Slovenia) or unlimitedly (eg in Germany, Sec 43 Copyright Act Germany, or Hungary, Sec 30 Copyright Act Hungary). A notable exception is Croatia, in which rights of use remain with the employee unless specified otherwise by law or contract (Sec 76 Copyright Act Croatia). All of these legislative rules are subject to differing contractual agreements. Jurisdictions that consider the employer as author include Ireland (Sec 23a Copyright Act Ireland) and the Netherlands (Sec 8 Copyright Act Netherlands).

105 Cf BGHZ 112, 243 = FCJ 27 September 1990 – I ZR 244/88 – GRUR 1991, 523, 525 – *Grabungsmaterialien*; Thomas Dreier, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 43 para 12; Haberstumpf (n 21) 825f; Peter W Heermann, 'Der Schutzumfang von Sprachwerken der Wissenschaft und die urheberrechtliche Stellung von Hochschulangehörigen' (1999) 6 GRUR 468, 474f.

106 Cf Dreier (n 105) para 12; Haberstumpf (n 21) 827; Heermann (n 105) 473.

107 Sec 8(3) *Zinātniskās darbības likums* 2005.On the definition of 'scientist', cf Sec 5(1), (3) of the Law.

108 Such as the German *Arbeitnehmererfindungsgesetz* 1957 or the Danish *Bekendtgørelse af lov om arbejdstageres opfindelser* 2012.

35 Yet, not all research data that is produced by academic professors necessarily falls within the scope of the privilege. In Germany, the privilege applies only to pure research.[109] Insofar as the creation of research data occurs at least partly in fulfilment of official duties, such as generating a patient file that is also used for research purposes, the employer is entitled to a right of use.[110] Where they generate research data within the scope of a certain commissioned research, such as in cooperation with a commercial company, researchers must generally also grant a right of use to their commissioners.[111]

36 If research data are collected on a larger scale and gathered in a research database, a *sui generis* database protection may exist. Generally, as previously established, the higher education institution or the commissioner or third-party funder is entitled to this protection. In this case, scientific freedom could make a reverse-direction impact: so as not to endanger the success of the research and to leave the decision of how and whether the data are used in future projects to the project-leading scientists, they should be granted a simple right of use to the *sui generis* database right.

2. Non-academic personnel

37 Especially in large-scale research projects, non-academic personnel may be deployed and commissioned with generating research data. If they do so, they also obtain possible rights thereto. This group of people includes, for example, medical technical assistants or student assistants. As their work is carried out fully bound by instructions, researchers' privileges stemming from the freedom of science do not apply.[112] Thereby, generally following from the employment contract, the employer obtains the exclusive rights to the protected materials. It has been proposed to grant the right of use to the research group leader, if the research results arose under their instruction.[113] However, since it is still the employer who has the right of direction, it appears reasonable to grant both—instructing researcher and employer—rights of use in these situations. This solution takes into account the economic interests of the employer and is in consistency with the database right on the one hand. While on the other hand, the further development of the research project is secured by granting the instructing researcher a right to use.[114]

3. Research assistants

38 Scientific freedom benefits not only academic professors: "every person acting or seeking to act scientifically is entitled to [it]".[115] This means that the employer is not granted a right of use to research data produced by research assistants in the scope of their own research, such as for academic qualifications.[116] For results of work carried out bound by instruction, however, the same principles as for non-academic personnel apply (cf supra, D.II.2.). This distinction can sometimes be difficult in larger research endeavors, such as in medical research. A variety of researchers of different hierarchies are regularly involved, working simultaneously on an overall project and on their own research as part of a sub-issue. In this case, origin and purpose of the specific research data are decisive: if it is material gathered while bound by instruction and supplied to the overall project, rights of use to it are granted. However, the employer does not obtain rights of use to research data gathered independently and texts (such as a dissertation) by the research assistant. With regard to the question of who obtains the rights of use, the same applies as for non-academic personnel (D.II.2.).

4. Externals (esp. external doctoral candidates)

39 Authors or rightholders that are not in employment or service relationships with an institution, such as external doctoral candidates and students, generally

109 Artur-Axel Wandtke, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), § 43 para 26.

110 FCJ 26 October 1951 – I ZR 93/51 – GRUR 1952, 257 – *Krankenhaus-Kartei*.

111 Ulrici, 'Kooperation in der Wissenschaft' (n 88) 328.

112 Cf BGHZ 112, 243 = FCJ 27 September 1990 – I ZR 244/88 – GRUR 1991, 523, 525 – *Grabungsmaterialien;* Haberstumpf (n 21) 827; Lauber-Rönsberg/Krahn/Baumann (n 90) 4.

113 Cf Haberstumpf (n 21) 827. Potentially with participation in a possible profit (cf Dreier (n 105) para 12; Haberstumpf (n 21) 827).

114 This functional approach can also be found in the decision BGHZ 112, 243 = FCJ 27 September 1990 – I ZR 244/88 – GRUR 1991, 523, 527 – *Grabungsmaterialien*.

115 BVerfGE 35, 79, 112 = German Federal Constitutional Court 29 May 1973 – 1 BvR 424/71 and 325/72 – NJW 1973, 1176, 1176.

116 Cf Dreier (n 105) para 12; Haberstumpf (n 21) 827; Heermann (n 105) 472; Lauber-Rönsberg/Krahn/Baumann (n 90) 4.

are not obliged to grant rights of use to the materials they create.[117] This is self-evident and unproblematic for texts produced for the purpose of academic qualification, such as dissertations. Often, however, these externals, particularly external doctoral candidates, are involved in generating research data for the collaborative project.[118] While they often do not receive compensation, they can access the research data pool and use it for their own research. Assuming that, nevertheless, these externals would not have to grant rights of use to the data they generate would lead to significant issues, eg, when the research database is intended to be made accessible to the public. It is possible to construe the supervision agreement as a *sui generis* contract, from which arises, *inter alia*, that the doctoral candidate grants rights of use to the research data produced by them for the overall project. This would presumably be compatible with the broad definition that the German FCJ gives of an employment relationship for the purposes of copyright, under which it suffices that the author "acts, in a more or less strongly dependent relationship, for the exploitation purposes of another", with the "intended purpose" of the work being decisive.[119] In favor of the external doctoral candidate, the contract could in turn give rise to protection and fiduciary duties of the supervising researcher. As such, it would for example have to be assured that the external doctoral candidate actually receives access to the data relevant for their research, especially when the supervisor changes institutions in the interim. Ideally, the parties' potential rights and obligations would be determined at the beginning of the cooperation. In doing so, for example, an arrangement would have to be made in case the supervision is terminated prematurely (in mutual agreement or otherwise).

5. Digression: Research cooperations

40 For research data created of research cooperations, the abovementioned generally applies, as well as potential explicit arrangements in research and development contracts. Despite the enormous scientific and economic relevance of research cooperations, jurisprudence and literature in this area are rare.[120] Proposals for a specific legal structure to be newly created for research cooperations are thus welcome.[121] Advantages of a cooperation structure with legal capacity would include the cooperation itself being the holder of the *sui generis* database right and concluding employment contracts in its own name.[122] Potential rights of use would fall directly to the research cooperation. Moreover, the cooperation structure could persist despite individual researchers withdrawing due to, eg, leaving the institution.[123] This way, access to the research data could be permanently ensured, benefiting the research project's success.[124] At the same time, concerns of scientific freedom related to the (continued) use of research data should be safeguarded via appropriate governance structures,[125] rather than battled out on the level of copyright rights of use.

E. International Research and Conflict of Laws: Aggravating the Problem

41 Particularly for international researchers, eg, visiting scholars, research fellows or visiting student researchers, the question of which country's copyright law applies to their research and their contribution to a larger research project can be both difficult to answer and decisive in determining which protection they receive and in what ways they can, in turn, use others' research data. The European conflict of laws rules apply where a court in a Member State is dealing with a case involving a conflict of laws, eg, where a German researcher working in a French institution claims an infringement of their research data by a third party located in the Netherlands. It

117 Cf Haberstumpf (n 21) 828; Heermann (n 105) 475.

118 On this, see also Ulrici, 'Kooperation in der Wissenschaft' (n 88) 147.

119 FCJ 22 February 1974 – I ZR 128/72 – GRUR 1974, 480, 482 – *Hummelrechte*.

120 Cf Möslein (n 91) 99. Cf also Nils Heide, 'Patentschutz und Patentlizenzen in Forschungskooperationen' (2013) InTer 2, who relates the lack of jurisprudence, besides the existing legal uncertainty, to the fact that possible disputes are more likely carried out before arbitral tribunals for reasons of secrecy.

121 Wolfram Eberbach, Peter Hommelhoff and Johannes Lappe, 'Eine Kooperationsform für die Wissenschaft' (2017) OdW 1, 5ff. See also Stefan J Geibel, 'Rechtsform und Zurechnungen zwischen Transparenz und Abschirmwirkung am Beispiel der Wissenschafts- und Forschungskooperationen' (2018) OdW 87; Möslein (n 91) 99.

122 Cf also Eberbach/Hommelhoff/Lappe (n 121) 8f.

123 Christoph Kumpan, 'Die Governance einer Forschungskooperationsgesellschaft – Struktur, Kompetenzen und Verfahren' (2018) OdW 115, 117.

124 Cf on this also BGHZ 112, 243 = FCJ 27 September 1990 – I ZR 244/88 – GRUR 1991, 523, 527 – *Grabungsmaterialien*.

125 See on this Kumpan (n 123) 117ff.

can lead to third countries' laws being applicable, even though they are not Member States (principle of universal application, Article 2 Rome I, Article 3 Rome II).

42 The law applicable to *infringements* of IP rights is determined by the conflicts rule of Article 8(1) Rome II, which follows the *lex loci protectionis* principle—the law of the country for which the plaintiff seeks protection is applicable.[126] The applicability of Article 8 Rome II is, however, debated in particular with regard to preliminary questions in copyright infringement proceedings. While Article 15 Rome II determines the scope of the laws applicable under the Regulation's conflicts rules, it is unclear whether this extends to the existence of copyright (and related rights) and its initial ownership.[127] The practical effect of denying applicability of Article 8 Rome II to these areas of copyright is that recourse to the conflict of laws rules of the competent Member States must be made. Often, Member States' respective conflicts rules will also apply the *lex loci protectionis* principle, so that identical results are achieved.[128] Yet, some Member States—most of which also follow the universality principle instead of the territoriality principle in questions of existence and initial ownership of copyright[129]—traditionally adhere to the rule of *lex originis* to questions of creation and initial ownership and would thus apply the law of the state in which the work was first made lawfully accessible to the public, or, if unpublished, the author's personal status.[130] The latest endeavor to provide unified conflict of laws rules for intellectual property, the Kyoto Guidelines, follows a third path and suggests that initial ownership in copyright and related rights should be governed by the law of the state with the closest connection to the creation of the work (cf Kyoto Guidelines 2020, Guideline 20(2)).[131]

43 For *contractual obligations* in connection with copyright, the general rules of Article 3, 4 Rome I apply.[132] These grant priority to the parties' choice of

126 Cf Rec 26 first sentence Rome II; ECJ, Case C-170/12 *Pinckney* [2013] ECLI:EU:C:2013:635. For the sake of completeness, note that for industrial property rights protected EU-wide (such as EU trade marks), Art 8(2) Rome II supersedes Art 8(1), which determines EU law, or, in case of gaps, the law of the place where the event which gave rise to the harm occurred, to be applicable.

127 Josef Drexl, *Münchener Kommentar zum BGB*, vol 13 (Franz Jürgen Säcker et al. eds, 8th edn, C.H. Beck 2021), Art 8 Rome II paras 177ff; Nerina Boschiero, 'Infringement of Intellectual Property Rights. A Commentary on Article 8 of the Rome II Regulation' (2007) 9 YPIL 87, 102f.

128 This is the case, eg, in Germany, Austria and France (cf for Germany, Regional Court of Munich I, Judgment of 14 May 2012, Case no. 21 O 14914/09, BeckRS 2012, 13691; for Austria, Austrian Supreme Court, Judgment of 17 December 2013, Case no. 4 Ob 184/13g, ZUM-RD 2014, 607, 610; for France, French Court of Cassation, Judgment of 10 April 2013, Case no. 11-12508, GRUR Int 2013, 955 (this judgment marks a shift in the French conflicts rule, which had previously applied the right of the country of origin)). On other Member States, see Boschiero (n 127) 99f; Toshiyuki Kono, 'Jurisdiction and Applicable Law in Matters of Intellectual Property' in Karen B Brown and David V Snyder (eds), *General Reports of the XVIIIth Congress of the International Academy of Comparative Law* (Springer 2012), 393, 410f; Pedro A de Miguel Asensio, 'The Private International Law of Intellectual Property and of Unfair Commercial Practices: Convergence or Divergence?' in Stefan Leible and Ansgar Ohly (eds), *Intellectual Property and International Private Law* (Mohr Siebeck 2009) 137 para 11. For completeness' sake, it should be noted that results may differ with respect to the nature of the conflicts rule – while Member States' autonomous conflicts rules may deem a state's entire law to be applicable and therefore allow *renvoi* (this is the case, eg, in Germany), Rome II excludes rules of private international law from referrals (cf Art 24 Rome II). Therefore, while the same state's law is applicable both under Rome II's and under Member States' conflicts rules, the latter includes the referred state's rules on conflict of laws, which may in turn provide for a different rule and therefore deem another law applicable (cf Drexl (n 127) para 176).

129 Michael Grünberger, 'Das Urheberrechtsstatut nach der Rom II-VO' 108 (2009) ZVglRWiss 134, 150.

130 These include Greece, Portugal and Romania (Art 67 *Νόμος Πνευματική Ιδιοκτησία* 1993 (Copyright Act Greece); Art 48(1) *Código civil* 1966 (Civil Code Portugal); Art 60 Romanian Private International Law Act); cf de Miguel Asensio, 'The Private International Law of Intellectual Property' (n 128) para 11; Katharina de la Durantaye, *Rome Regulations Commentary* (Gralf-Peter Calliess and Moritz Renner eds, 3rd edn, Wolters Kluwer 2020), Art 8 Rome II para 3.

131 This is assumed to be the state in which the person who created the subject matter of the work was habitually resident at the time of creation. For the existence, scope and transferability of IP rights, as well as their infringement, the Kyoto Guidelines follow the *lex loci protectionis* principle (cf Kyoto Guidelines, Guidelines 19, 25.).

132 Employment contracts with research institutions will usually fall within the scope of Art 8 Rome I. For researchers in public service relationships (such as state university professors), private law (and thus, Rome I) applies insofar as they do not exercise sovereign powers. Cf Dieter Martiny, ***Münchener Kommentar zum BGB***, vol 13 (Franz Jürgen Säcker et al. eds, 8th edn, C.H. Beck 2021), Art 1 Rome I para 6; Peter Mankowski, *Europäisches Zivilprozess- und Kollisionsrecht*, vol 1 (Thomas Rauscher ed, 5th edn, Otto Schmidt 2021), Art 20 Brussels Ia Regulation, para 79ff; Ulrich Magnus, *Internationales Vertragsrecht 1* (Julius von Staudinger ed, Sellier de Gruyter 2021), Art 8 Rome I para 47.

law. If no choice of law was made,[133] they alternatively provide for connections based on the type of contract (Article 4(1) Rome I). Yet, contracts dealing with copyright, such as licensing agreements or assignments, may fall under multiple contract types and therefore be difficult to categorize.[134] In the absence of a specific contract type, the law of the country "where the party required to effect the characteristic performance of the contract has his habitual residence"[135] will be applicable under Article 4(2) Rome I. However, establishing the characteristic performance in contracts related to copyright can be equally difficult due to their wide variety and differing levels of complexity.[136] While it appears evident that the author's or rightholder's performance is characteristic when it is given in exchange for financial remuneration, the determination is less clear when the other party itself is obligated to perform specific actions, such as to exploit a work or exercise rights granted to it. The situation becomes even more complicated if one contract is concluded between multiple parties (granting, for example, exploitation rights to multiple persons to use the work or when multiple authors assign their rights to another party). Often, therefore, the law applicable to IP contracts will be determined on a case-by-case basis under Article 4(4) or under Article 4(3), as the law of the country with which the contract is most closely connected.[137] This could be the country of the author or copyright holder, the country of the other party (eg, licensee or assignee), the country for which protection exists, or another country depending on the specifics of the contract. Similarly, the Kyoto Guidelines provide that, in the absence of a choice of law by the parties, contracts dealing with IP granted for more than one state (other than employment contracts) are governed by the law of the state with which the contract is most closely connected (Kyoto Guidelines, Guideline 22(2)). Connecting factors include the common habitual residence of the parties, the habitual residence of the party effecting the characteristic performance,[138] and the habitual residence of one of the parties when this habitual residence is located in one of the states covered by the contract. Employment contracts in employment relationships where employees may create intellectual property should be governed, in the absence of a choice of law by the parties, by the law of the state in which or from which the employee habitually carries out their work in performance of the contract (Kyoto Guidelines, Guideline 23(3)).

44 In international research projects, where a large number of researchers collaborate in creating research data that may be protected by copyright or related rights, Articles 4(1) and 4(2) Rome I cannot provide satisfactory results. The closest connection must therefore be determined under Article 4(4). We submit that the countries of the collaborating researchers cannot provide the closest connection, because this would result in different countries' laws applying simultaneously. The law of the country in which the research is conducted (eg, if there is one research institution) appears to be better suited—it falls short, however, if multiple institutions collaborate and research is conducted multi-nationally. In this case, there may be need to identify the main seat of a research project, such as the leading institution.

F. Access and Usability: Current Practices and Future Solutions

45 The practical need for access to research and research data goes far beyond its current accessibility.

133 A choice of law can also be implied by other terms of contract, cf Art 3(1) second sentence Rome I. On requirements to assume implied choice of law, see Dieter Martiny, *Münchener Kommentar zum BGB*, vol 13 (Franz Jürgen Säcker et al. eds, 8th edn, C.H. Beck 2021), Art 3 Rome I paras 46ff; Richard Plender and Michael Wilderspin, *The European Private International Law of Obligations* (5th edn, Sweet & Maxwell 2020, paras 6-026ff.

134 Kono (n 128) 410f; Paul LC Torremans, 'Licences and Assignments of Intellectual Property Rights under the Rome I Regulation' (2008) 4 JPIL 397, 403; Pedro A de Miguel Asensio, 'Applicable Law in the Absence of Choice to Contracts Relating to Intellectual or Industrial Property Rights' (2008) 10 YPIL 199, 207ff, examining a number of specific IP contracts to determine whether and where they fall with regards to Art 4(1) Rome I.

135 The habitual residence of natural persons is their center of life, which requires residence for a certain time. If they are acting within the scope of their business activity, which includes dependent employment, their principal place of business is decisive. International researchers will generally only be with an institution for a finite period of time with the intention of returning to their home country or relocating elsewhere after the research has been completed. This *animus revertendi* has the effect of applying the law of the researcher's home country, rather than the law of the country in which the researcher is currently located, cf Georg John, 'Der Begriff des gewöhnlichen Aufenthaltes und seine Bedeutung im europäischen Privat- und Zivilverfahrensrecht (Teil I)' (2018) 2 GPR 70, 78; Marc-Philippe Weller and Alix Schulz, 'Unterhaltsklage nach Kindesentführung: Zuständigkeit am „unrechtmäßigen" gewöhnlichen Aufenthalt des Kindes?' (2015) 2 IPRax 176, 179f.

136 Kono (n 128) 410f; Torremans (n 134) 403f; de Miguel Asensio, 'Applicable Law in the Absence of Choice' (n 134) 207ff.

137 Torremans (n 134) 404.

138 The Guidelines acknowledge that, in case of complex IP contracts, it is not always possible to identify a characteristic performance (cf Kyoto Guidelines, 52, para 41).

This is illustrated by the advent of so-called 'shadow libraries', most famously Sci-Hub, whose self-proclaimed aim is "to provide free and unrestricted access to all scientific knowledge" and which boasts a collection of over 88 million pdf files.[139] The website violates copyright and related rights by bypassing access limitations of established scientific websites, allowing users to access existing research at no cost. Other initiatives, like the Internet Archive,[140] merely act as digital archives of cultural artifacts such as books, images or audio recordings, also including internet sites themselves. Through its "Open Library", the Internet Archive provides free downloads of public domain works and digital lending of modern books. A third type of platform, like Researchgate or SSRN, operates more like a repository rather than a library and enables researchers themselves to share their publications, if they so wish. It is highly disputed whether these business models are in compliance with the rules of copyright law.[141] However, the popularity of such websites (Sci-Hub claims that in June 2022, over 10 million papers were downloaded around the world; at the time of completion of this article, nearly 2 million papers had been downloaded via SSRN in the last 30 days)[142] evidences a practical need to facilitate low-threshold access to scientific knowledge.

46 Currently, research data can be used lawfully either where such use is allowed under a statutory exception or limitation or where the authors themselves have permitted such use factually or contractually. The second option is relevant in particular for research conducted at universities and some independent research institutions, which adopt so-called open science policies. In keeping with the EU's legislative trend, future rights of use tailored to research data are also conceivable.

139 Cf <https://sci-hub.hkvisa.net/> accessed 5 July 2022; <https://sci-hub.ru/about> accessed 5 July 2022.

140 Cf <https://archive.org/about/> accessed 5 July 2022.

141 The Internet Archive was sued in the State of New York by four major US publishers (Hachette, Harper Collins, Wiley and Penguin Random House), claiming that its "Controlled Digital Lending" program infringes the publishers' copyright and is not covered by fair use or the first sale doctrine (Hachette Book Group, Inc. et al. v. Internet Archive, Case No. 1:20-CV-04160-JGK); Researchgate was sued in Germany by several publishers from the "Coalition for Responsible Sharing", led by Elsevier and the American Chemical Society, for making available several publications on the platform (LG München I, 31.1.2022 – 21 O 14450/17).

142 Cf <https://sci-hub.ru/stats> accessed 5 July 2022; <https://papers.ssrn.com/> accessed 12 July 2022.

I. Statutory exceptions and limitations

47 Perhaps most importantly, a number of uses of research data are permitted by law without needing to acquire a license from the rightholder. While such statutory exceptions do not apply specifically to research data, they include certain of its uses within their broader scope. If the data are used for scientific research or educational activities, the mandatory exception for text and data mining for the purposes of scientific research (Article 3 DSM Directive)[143] and the facultative exceptions for scientific research (Article 5(3) lit a InfoSoc Directive)[144] can apply. Researchers can rely on these exceptions both when using and archiving pre-existing works in a research database, as well as using research data generated by others (within certain boundaries). Further, the right of quotation provided in Article 5(3) lit d InfoSoc Directive can enable not only the collection of preexisting research data itself, but also the use of these data within one's own scientific research.

48 The exception for scientific research applies to the methodical pursuit of knowledge in a broad sense, including (but not limited to) research conducted by university professors, research institutions and research assistants.[145] The exception's scope is determined by the Member States when transposing the InfoSoc Directive into national law.[146] According to Section 60c(1) of the German Copyright Act, "up to 15 per cent of a work [or an object protected by related rights] may be reproduced, distributed and made available to the public for the purpose of non-commercial scientific research", for a specifically limited circle of persons for their personal scientific research (number 1) or for individual third persons

143 Romain Meys, 'Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market: Are European Database Protection Rules Still Threatening the Development of Artificial Intelligence?' (2020) 5 GRUR Int. 457, 465.

144 Frederik Leenen, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), InfoSoc Directive Art 5 para 101.

145 Thomas Dreier, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 60c para 1.

146 Art 5(3) lit a InfoSoc Directive stipulates only certain minimum requirements: The work must be used for the sole purpose of scientific research; the source, including the author's name, should be indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved. On Member State's implementation, cf Sec 19(2) Copyright Act Estonia; Sec 60c Copyright Act Germany; Secs 22, 58(5) Copyright Act Lithuania; Sec 9(1) lit h Copyright Act Malta; Sec 44 *Autorský Zákon* 2015 (Copyright Act Slovakia).

insofar as this serves the monitoring of the quality of scientific research (number 2). For example, a group of researchers can set up a shared database in which (up to 15 per cent of) sections of relevant monographs are made available. Also, "illustrations, isolated articles from the same professional or scientific journal, other small-scale works and out-of-commerce works" may be used fully (paragraph 3).[147] This means that pre-existing works can also be included in a research database (within the permitted scope). However, this research database can then only be made available to a personally distinct group of researchers or for the purposes of monitoring the results of the research by third persons.

49 Section 60c(2) of the German Copyright Act allows the reproduction of a work on a larger scale (75 per cent), but only for personal scientific research, meaning the used extracts of works cannot be made available to others within a research database.

50 Not only do researchers deal with individual works and related rights, they also use text and data mining to automatically search and analyze large quantities of text and data.[148] Although text and data mining as such, ie, the automated evaluation alone, is not relevant to copyright law,[149] it requires data to be in a machine-readable format (corpus), which generally necessitates a reproduction of the material.[150] Moreover, in the context of machine learning algorithms it might be necessary to annotate and adjust the text or data before using it for the training of the system.[151] The DSM Directive establishes a uniform European foundation for text and data mining for the purposes of scientific research (Article 3 DSM Directive) and allows it, in limited scope, for other purposes (Article 4 DSM Directive). In Germany, it is implemented in Section 60d Copyright Act (revised), and in Austria in Section 42h Copyright Act. Section 60d of the German Copyright Act permits both potential reproductions of the source material (paragraph 2, first sentence), as well as making the corpus available to the public for a specifically limited circle of persons for their joint scientific research, or to individual third persons for the purpose of monitoring the quality of scientific research (paragraph 4, first sentence). This exception, however, applies only to material to which a lawful access already exists; it does not create a "claim for access to protected source material".[152] The revised provision also allows for the material reproduced in the area of scientific research to be permanently stored and retained (Article 3(2) DSM Directive, Section 60d(5) Copyright Act Germany).

51 The exceptions both in the DSM Directive and the InfoSoc Directive privilege non-commercial research (Article 2(1) lit a var 1 DSM Directive, Article 5(3) lit a InfoSoc Directive). The non-commercial character of a research project is not forfeited by third-party funding or by the prospect of a profitable publication.[153] The DSM Directive also privileges research with commercial gains, insofar as potential profits are fully reinvested into the research (Article 2(1) lit a var 2 DSM Directive).[154]

52 Finally, the right of quotation can allow the use of research data protected as works or under related rights within one's own research, even if this research does not pass the originality threshold to warrant copyright protection itself.[155] To fall within the right of quotation, the data must be used to illustrate an assertion, defend an opinion or allow an intellectual comparison between the data and the assertions of its user.[156] The requirements are context-specific, meaning they are determined on a case-by-case basis by considering the specific use at hand. For quotations of text, the user must "establish a direct and close link between the quoted

147 For articles from non-scientific journals, however, the per cent boundary of para 1 (or para 2) applies.

148 BT-Drs 18/12329, 40. On the scientific relevance of text and data mining, see only Benjamin Raue, 'Rechtssicherheit für datengestützte Forschung' (2019) 8/9 ZUM 684; Louisa Specht, 'Die neue Schrankenregelung für Text und Data Mining und ihre Bedeutung für die Wissenschaft' (2018) OdW 285.

149 Cf BT-Drs 18/12329, 40; Rec 9 DSM Directive.

150 Cf Katharina de la Durantaye, 'Neues Urheberrecht für Bildung und Wissenschaft. Eine kritische Würdigung des Gesetzentwurfs' (2017) 6 GRUR 558, 561; Raue (n 148) 685.

151 Cf Lisa Käde, *Kreative Maschinen und Urheberrecht* (Nomos 2021) 65ff, 70f; Björn Steinrötter and Lina-Marie Schauer, 'Text und Data Mining, Forschung und Lehre' in Malek Barudi (ed), *Das neue Urheberrecht* (1st edn, Nomos 2021) para 5; Philipp Hacker, 'Computer-Generated Works im deutschen Urheberrecht? Überlegungen zur Schutzfähigkeit von KI-Erzeugnissen in komplexen technischen Entwicklungsprozessen' in Linda Kuschel, Sven Asmussen and Sebastian Golla (eds), *Intelligente Systeme – Intelligentes Recht* (Nomos 2021) 234f.

152 BT-Drs 18/12329, 41.

153 BT-Drs 18/12329, 39. Cf also Thomas Dreier, *Urheberrechtsgesetz* (Thomas Dreier and Gernot Schulze eds, 7th edn, C.H. Beck 2022), § 60c para 6; Stefan Lüft, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), § 60c para 12.

154 Cf on this also Raue (n 148) 690.

155 ECJ, Case C-145/10 Painer [2011] ECLI:EU:C:2011:798, para 137.

156 ECJ, Case C-516/17 Spiegel Online [2019] ECLI:EU:C:2019:625, para 78.

work and his own reflections, thereby allowing for an intellectual comparison to be made with the work of another" and "the use of the quoted work must be secondary in relation to the assertions of that user".[157] This exception is voluntary for the Member States.[158] A particular limitation for research data is that the exception requires the cited works to have previously been lawfully made available to the public by the rightholder.[159] This may not always be the case, particularly in a natural sciences context, as research data may often be shared only between peers, without satisfying the requirements for being made available to the public (ie, allowing access by an indeterminate number of potential recipients and involving a fairly large number of people).[160]

II. Open Science

53 Besides the statutory exceptions for the handling of research data, which may be unclear and/or too restrictive in individual cases, a rising number of researchers subscribe to so-called open science policies to allow the exploitation of their research and research data in a controlled form. Most higher education institutions have introduced open access and open science policies, which require a portion of or even all research conducted under the aegis of the institution to be accessible freely, ie publicly and free of charge, via the internet (eg in open access research journals and/or institutional or disciplinary digital repositories).[161] This is also the case for many prestigious research funding organizations or programs, such as the European Research Council's funding under the Horizon Europe program.[162] Where open science policies exist, free accessibility extends to research data as such, as well as research software and teaching materials. These policies are intended to benefit the free dissemination of knowledge and foster good scientific practice, as well as reduce the cost of scientific publication.

54 Depending on the policy, primary or secondary open access publication is required.[163] In case of secondary open access publication, ie after the research (data) has already been published in a periodical scientific journal,[164] Germany grants a digital second publication right (Section 38(4) Copyright Act Germany), which gives authors the right to republish the accepted manuscript of their work 12 months after first publication. This right requires the work to have been created within the course of research financed at least in half by public funding and to have been published in a periodical collection (appearing at least bi-annually). The republication cannot serve a commercial purpose. This right cannot be excluded in the contract between author and (first) publisher. While this right is an important initiative on the way to more open access-friendly legislation, it is arguably too narrow to be truly effective. Particularly in the natural sciences, the waiting period of 12 months may cause the work to lose relevance before being available for republishing, the manuscript version is not ideal to encourage academic discussion as citations are made difficult (the relevant journal page numbers are not available), and the exception imposes artificial requirements excluding a large number of research papers ab initio.

55 In 2015, a German university has amended its bylaws to make secondary open access publication a mandatory obligation for its researchers; this has been challenged and is currently under judicial review with the German Federal Constitutional Court.[165] If there is no institutional policy in place,

157 ECJ, Case C-516/17 Spiegel Online [2019] ECLI:EU:C:2019:625, para 79.

158 Note that the right of quotation has been made mandatory in the context of user generated content on online content-sharing service platforms by Art 17(7)(2) lit a DSM Directive.

159 ECJ, Case C-145/10 Painer [2011] ECLI:EU:C:2011:798, para 127; ECJ, Case C-516/17 Spiegel Online [2019] ECLI:EU:C:2019:625, para 89.

160 ECJ, Case C-392/19 VG Bild-Kunst [2021] ECLI:EU:C:2021:181, para 31; ECJ, Case C-263/18 Tom Kabinet [2019] EU:C:2019:1111, para 66; ECJ, Case C-265/16 VCAST [2017] EU:C:2017:913, para 45.

161 Cf on the definition of open access, Budapest Open Access Initiative, Declaration (2002) <https://www.budapestopenaccessinitiative.org/read/> accessed 19 September 2022.

162 Horizon Europe, the EU's research and innovation funding programme from 2021-2027, requires all scientific publications to be open access and research data management under the FAIR Principles. Cf European Research Council, *Open Research Data and Data Management Plans – Information for ERC grantees*, 2022, p 4.

163 Discussing the constitutional permissibility of publication obligations in funding eligibility conditions in depth, Fehling (n 2) 179.

164 As opposed to scientific publication series, handbooks, monographs, commentaries and similar singular publications (Artur-Axel Wandtke and Eva-Marie König, *Praxiskommentar Urheberrecht* (Artur-Axel Wandtke and Winfried Bullinger eds, 5th edn, C.H. Beck 2019), § 38 para 20), but also online repositories such as JSTOR or SSRN.

165 The case is on file with the Higher Administrative Court of Baden-Württemberg, which in 2016 suspended the proceedings to request a ruling from the Federal Constitutional Court (Case

the researcher as copyright holder can decide whether they wish to publish open access or not; however, university target obligations or financing incentives may influence this decision.[166] The research (data) so published may be used lawfully as determined by the repository rules; usually, there will be no legal limitations beyond authors retaining control over the integrity of their work and their proper acknowledgement and citation.[167] This is often achieved by employing Creative Commons licenses generally or Open Data Commons licenses specifically for data collections.

56 Not all research data can or should be made openly available. However, where open access is not provided to research data, they should at least be processed in a sustainable way, ensuring access and reusability by others. One way to arrive at this goal is by implementing the "FAIR Data Principles" (Findable, Accessible, Interoperable and Reusable).[168]

III. Current developments and future desiderata

57 Currently, a number of endeavors at the EU level target an improvement of access to data and data governance. For instance, the Data Act proposal published in February of this year clarifies the lack of protection of machine-generated data under the *sui generis* database right, which had thus far been subject to legal uncertainty.[169] Further, the proposal for a Data Governance Act aims to enable the reuse of public-sector data subject to the rights of others. Although the Act does not apply to IP rights, public-sector bodies are encouraged to exercise their copyright in a way that facilitates re-use.[170] This may point to public sector-conducted or -financed research using open access policies on a larger scale in the future. In 2019, the recast Open Data Directive had already stressed the importance of open data licensing for public sector data.[171] Finally, the EU's 2020 IP Action Plan recognizes researchers' struggle with IP protection, promising to boost IP asset management by increasing know-how and to "[take] steps [...] to ensure that publicly funded IP is used in a fair and effective manner".[172]

58 But individual organizations are also becoming more aware of the benefits of research data management. Research institutions increasingly provide model contracts or templates containing clear provisions on rights to research data and individual project agreements are becoming more common. Nevertheless, this is by no means standard practice for research projects and should be continually encouraged at the institutional level.

59 Adjacent to these piecemeal developments largely resting on recommendations and organizations' own actions, there is a case to be made for creating clear and universal rules on IP protection for research data. Of course, such legislation would need to respect fundamental rights, specifically fall within the limits of the freedom of science, property rights and the freedom of business (particularly of scientific publishers).[173] The abovementioned Latvian Law on Scientific Activity could well serve as a model for EU-level legislation, as it gives clear definitions for research workers and sets out unambiguous rules for intellectual property ownership to research.

---

no. 9 S 2056/16), cf on this Manfred Löwisch, 'Streit um die Zweitveröffentlichungspflicht geht zum Bundesverfassungsgericht' (2018) OdW 43. To date, the Constitutional Court has not ruled. Generally on secondary publication obligations in higher education bylaws, see Volker M Haug, 'Open Access in Baden-Württemberg: Rechtswidriger Zweitveröffentlichungszwang zwischen Urheber- und Hochschulrecht' (2019) OdW 89.

166 When evaluating research proposals for Horizon Europe grants, the quality and appropriateness of open science practices is taken into account. Cf European Commission, Directorate-General for Research and Innovation, *Horizon Europe, Open Science: Early Knowledge and Data Sharing, and Open Collaboration* (2021) https://data.europa.eu/doi/10.2777/79699 accessed 19 September 2022.

167 Budapest Open Access Initiative, Declaration (2002) (n 161). For Open Data, cf the Open Knowledge Foundation's Open Definition <http://opendefinition.org/> accessed 19 September 2022.

168 FAIR Guiding Principles for scientific data management and stewardship, <www.go-fair.org/fair-principles/> accessed 19 September 2022.

169 European Commission, Data Act Proposal, 23 February 2022, COM(2022) 68 final, p 5, Rec 84, Art 35.

170 European Council, Council Approved Data Governance Act Proposal, 4 May 2022, PE-CONS 85/21, Rec 17, 18.

171 Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, Rec 44.

172 European Commission, IP Action Plan, 25 November 2020, COM(2020) 760 final.

173 For an in-depth assessment under German law, see Philipp Overkamp and Miriam Tormin, 'Staatliche Steuerungsmöglichkeiten zur Förderung des Teilens von Forschungsdaten', (2022) OdW 39.

G. Conclusion

60 There are many ways in which research data can be protected under EU copyright law, either as copyrighted works or via related rights. Even in areas where the law is not harmonized, parallel protection regimes can often be found in the Member States. This protection can impede the access to and use of research data, particularly in the context of larger research groups as well as in international and multi-organizational research projects. These complexities could be reduced by introducing legal instruments that take into account the particularities of research activity and research data, thereby providing a more functional approach to copyright in this area. The Latvian Law on Scientific Activity is a lighthouse in this regard, as it gives clear definitions for researchers and sets out unambiguous rules for intellectual property ownership of research. The current momentum of data-related regulation on the EU level could well be used to further this aim.

61 In the predominant absence of specific rules for access to and use of research data in the EU and its Member States, it is crucial that researchers themselves negotiate contractual rules to govern their legal relationships. The protection copyright offers for research data proves useful only where it is actively wielded, rather than subsequently applied. Conducting research together with other researchers, assistants and non-academic personnel without individual project agreements may create a thicket of rights that can jeopardize the success of the project. Researchers should therefore take care to negotiate the rights to their data, as well as who and how it can be used in advance within the legal framework provided. Research institutions should make it their practice to provide guidelines and complete detailed contractual rules on research data with their students and personnel, in order to minimize legal uncertainty and ensure the copyright regime does not become an encumbrance for future developments.