

## EARE's Position on the European Data Union Strategy

The European Alliance for Research Excellence (EARE) welcomes the publication of the European Data Union strategy, as an important step towards opening data and unlocking its potential for research and innovation in Europe. This strategy can position the EU as a global leader in research and innovation, provided it addresses persistent barriers that limit data access, and use of data by researchers and innovators.

### **A positive step to boost data access for researchers and innovators**

EARE supports the strategy's objective to scale up access to data for researchers and innovators for AI development. As stressed in EARE's [contribution](#) to the European Data Union strategy, large, and diverse datasets are the foundation for AI development. Access to this data prevents biases and allows researchers and innovators to develop European AI models that can be able to compete globally.

In this context, EARE welcomes the proposal for the creation of data labs, which will provide companies and researchers with high-quality datasets for AI development. The effectiveness of these data labs will depend on their practical accessibility. Clear access conditions and proportionate governance are needed to ensure they are usable by smaller research actors, startups, and spin-offs, and do not become closed or overly complex structures.

Beyond availability, data access is also hindered by regulatory fragmentation. Simplifying and consolidating the EU data framework is essential to provide researchers, innovators and startups with legal clarity and workable access to data across the Single Market.

### **Additional barriers to publicly funded research data must be addressed**

While the strategy emphasizes boosting data access for AI, EARE believes unlocking publicly funded research and has been financed by taxpayers is critical. Although the strategy acknowledges that the upcoming European Research Area Act will strengthen the legal conditions to share, access and reuse of publicly funded research results, EARE calls for additional actions. Specifically, the introduction of secondary publication rights (SPRs) at EU level is essential. Secondary publication rights would allow researchers to publish their work in open repositories after a short embargo, enhancing transparency, reproducibility and dataset quality.

## **Strengthening Text and Data Mining (TDM) exceptions is also essential to unlock the full potential of data in the EU**

The Data Union strategy also overlooks a critical enabler of research and innovation: text and data mining (TDM). TDM is essential for researchers, innovators, libraries, and start-ups to analyse large datasets and develop accurate AI models. For example, the University of Oslo, leveraged TDM to process thousands of texts to discover systematic biases on exclusions of minorities, women, and lower classes, which served to rewrite cultural history with greater accuracy. Without strong TDM rights, Europe would risk limiting access to data and slowing progress in AI development. Legal uncertainty around TDM also affects startups, scale-ups, and research spin-offs, for whom unclear exceptions and restrictive licensing translate into higher costs, delayed innovation and reduced incentives to invest in data-driven research in Europe.

Articles 3 and 4 of the EU Copyright Directive provide essential TDM exceptions, intended to enable research without the need for prior authorisation. However, their effectiveness is being undermined by the growing push for licensing schemes. A [survey](#) of researchers conducted as part of the European Commission's study on improving access to and reuse of research results found that many researchers refrain from using TDM tools despite the existence of the exceptions, primarily due to fear of copyright infringement, uncertainty about lawful access, and concerns about breaching licence terms. Even when TDM would be lawful, researchers often face licences that limit TDM exceptions, turn freely usable content into restricted material, reduce data availability and diversity, increase bias in AI models, and ultimately limit research activity and innovation.

In this context, it is concerning that in the [staff working document](#) accompanying the strategy, a survey of stakeholders accused libraries of breaking the law by allowing legal deposit content to be used for the training of AI large language models. This narrative risks discouraging legitimate research practices. Libraries across Europe, such as the [National Library of Norway](#), are exploring the potential use of AI in libraries to improve access to knowledge. For these efforts to succeed, Article 3 of the Copyright Directive on TDM exceptions is essential, as it provides the legal foundation for using data for scientific research.

## **EU's data strategy should reflect the cross-sector and cross-border nature of research and innovation landscape**

Today's research and innovation reality usually involves public and private participation and cross-border cooperation among Member States. While the Data Union strategy recognizes the importance of private-public collaboration to unlock the potential of data, it overlooks

the restriction of TDM exceptions to non-commercial use. This limitation forces researchers and innovators into complex licensing arrangements if their work has potential commercial implications. It creates significant legal uncertainty, particularly in public-private partnerships or when non-commercial research leads to commercial applications, a common scenario in sectors such as health. These constraints are further intensified by fragmented legal frameworks across Member States, which hinder cross-border collaboration and increase burdens for researchers and innovators. This situation discourages researchers and innovators from using data, fearing copyright infringement. The Data Union strategy falls short of addressing this issue, leaving an obstacle to unlock Europe's data potential.

### **The way forward to unlock EU's data potential**

Moving forward, the EU should build on the progress made through the European Data Strategy, the AI Act, and the AI Continent Plan. EARE calls on upcoming initiatives such as the European Research Act (ERA), the European Innovation Act, and the review of the Copyright Directive to strengthen TDM exceptions and remove legal and technical barriers to data access for researchers and innovators. Only by removing barriers to data can the EU unlock the full potential of data-driven research and innovation to compete globally.



**About EARE:** The European Alliance for Research Excellence (EARE) was convened by Microsoft in 2017, and now brings together nine members from the research and innovation ecosystem in Europe, including the Association of European Research Libraries (LIBER Europe), the European Bureau of Library, Information and Documentation Associations (EBLIDA), BSA | The Software Alliance, Microsoft, Allied for Startups, LACA, Research Libraries UK, SCONUL (Society of College, National and University Libraries), and UCL (University College London) Library, advocating for the EU to live up to its innovation potential in the digital economy.