LIBER (Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries) is the voice of Europe’s research library community.

Some 450 national, university and other libraries are part of LIBER and our wider network includes goal-oriented partnerships with other organisations in Europe and beyond. For nearly 50 years, we have worked together to meet our mission of enabling world class research in Europe.

We have worked consistently on the issue of text and data mining since 2013 participating in the TDM working group of Licences for Europe, launching the Hague Declaration on Knowledge Discovery in the Digital Age in 2015. The Hague Declaration called for immediate changes to intellectual property law and was signed by nearly 900 individuals and institutions. We have also been active working with Council and MEPs as part of the Digital Single Market Directive.

Broadly we welcome the German draft on Text and Data Mining but would like to raise the following issues:

1. We think it is important that the implementation supports AI and machine learning and not just the related but specific field of data mining. We therefore would urge that more general terms such as data analytics and machine learning are used in the text. We also believe that some of the terminology used in the draft is specific to data mining rather than machine learning in general. (see 3 below)

2. Article 3 of the Digital Single Market Directive allows commercial data mining by public interest organisations subject to certain stipulations as correctly outlined in § 60d (2)3. However § 60d (3) 1 and 2 appear to undermine and conflict with the directive by limiting absolutely commercial activity in the case of 3, and being silent in subsection 2. We believe this section could be reworded and closer reflect the directive which only limits commercial activity when there is controlling influence by a commercial entity. This is very important for the knowledge transfer agenda in universities, as well as public private partnerships that take place amongst cultural heritage organisations[9W1].

3. We strongly welcome the ability as outlined in section 4 of § 60d to share the results of machine learning however believe that the making available is for more than just “checking purposes” – it is for research, evaluation, model training, algorithm testing, writing, improving algorithms, machine learning etc. We are also concerned that the last sentence of section 4 means that the valuable data can no longer be shared in future projects that could use the work of the initial project for which the data was first created. Research is building on the work of those that have gone before, and we believe that the wording needs to reflect that science comes from standing on the shoulders of giants.

4. In line with the previous comment we also wish to raise the issue that the draft is silent on how to share the results of the data mining. Researchers need to be able to
share the data and the results of machine learning, and we recommend that the German government further reflects on the sharing aspects of machine learning and introduces an appropriate exception allowing communication to the public for scientific research purposes in line with 5.3a of the Information Society Directive.

5. We strongly welcome section 5, which reflects the data management plans of Deutsche Forschungsgemeinschaft and leaves it up to the researcher how and where to store the data securely. However Paragraph 5 should also make it clear that:
   i) the authorised institutions may also make additional copies and processing required for storage and long-term archiving;
   ii) the data may be made available again for future research purposes in line with the law.

6. Conversely we are very concerned that section 6 underlines the right of publishers to use technical protection measures, and yet is not balanced at all with any measures for those able to enjoy this exception to quickly access the data they are being locked out from, and for compensation to be awarded when this is unacceptably long. LIBER is currently collecting data on how long universities are being blocked from access to material they are handsomely paying publishers for and it can be many months. This contrasts with consumers who will get very swift resolution to any music tracks they are blocked from accessing on platforms such as iTunes and Amazon. Given the public interest, and the billions of euros being invested in research and acquisition budgets we believe it is incumbent on the German government to introduce new and swift measures to allow the circumvention of technical protection measures.

7. We also raise the issue of the need to digitise analogue materials in order to undertake data mining which we believe is covered in the wording. We would also ask the drafters to reflect on whether the current wording related to “making available” is sufficient to allow remote machine learning, so researchers can undertake data mining within their own workflow rather than having to physically relocate to the library.

8. Finally in § 44b (3) we are concerned as to how this drafting may be interpreted as anything on a computer screen is theoretically “machine readable” – a PDF, terms and conditions etc. It is vitally important that it is clear this relates to widely used machine readable “standards” such as robots.txt protocol. If this is not the case then anything is machine readable, and the wording is tantamount to requiring all terms and conditions on a website having to be read and interpreted by a human one by one. This was not the intention of the European Council or the European Parliament who were expecting the adoption of machine readable standards to be used.

Please note that this has been prepared using translation technology so some of the nuance in the original German may have been lost.