Moreover, other services are exploring the potential of social networks as a vehicle to gather and disseminate research. Mendeley is a popular reference manager software produced by Elsevier that allows sharing research papers and online collaboration among over eight million researchers. Another network, colwiz (collective wizdom), launched in 2011 and provided interactive digital collaboration and free reference management services for researchers in academia, industry and government globally. They also developed the ACS Chemwork platform for the American Chemical Society. In 2013, Taylor & Francis incorporated colwiz's interactive PDF reader into their journals platform and in 2017 its parent, Informa, acquired the whole company. In 2016, the company also developed the wizdom.ai research intelligence product (see section 4.7 *Tools, apps and new services for funders and institutions*). At the time of writing, colwiz functionality was being merged into wizdom.ai to develop an intelligent research assistant under the wizdom.ai brand.

Bibliography management software (such as Endnote (Thomson Reuters), Flow (Proquest), Pages (Springer), Zotero, etc.) also allows users to share their research libraries with other users but typically the sharing is inherently one-to-one or one-to-few, or restrictions on the numbers of users with whom content may be shared are explicitly enforced

The popularity of SCNs is perhaps an indication of the way in which authors prefer to share their articles. However - unlike open access repositories - academic social networks do not routinely check for copyright compliance, and therefore much of their content is illegally posted and hosted (Jamali, 2017).

Uncertainty over the copyright status of academic papers hosted on social networking sites raises concerns over the persistence of such content (Chawla, 2017) and the ethics of ASN services themselves (Fortney & Gonder, 2015). It seems likely that ECRs at any rate are less likely to make their research publications available for immediate downloads rather than invite an invitation to share.

4.8 Text and data mining

Text and data mining (TDM) has the potential to transform the way scientists use the literature (Nature 2012). It is expected to grow in importance, driven by greater availability of digital corpuses, increasing computer capabilities and easier-to-use software, and wider access to content. The Publishing Research Consortium report Text Mining and Scholarly Publishing (Clark 2013) gives a good introduction to TDM (see also Johnson, Fernholz and Fosci, 2016; Clark, Jensen, & Campbell, 2014; and Smit & van der Graaf, 2011).

TDM draws on natural language processing and information extraction to identify patterns and find new knowledge from collections of textual content. Semantic enrichment and tagging of content are likely to enhance TDM capabilities. At present TDM is most common in life sciences research, in particular within pharmaceutical companies, but relatively little used elsewhere.

The main challenges for more widespread adoption are legal uncertainties as to what is permitted, and the lack of an efficient licensing regime (see *2.16.8 Text and data mining rights*); technical issues such as standard content formats including basic common ontologies; the need for content aggregation to permit mining cross-publisher corpuses; the costs and technical skills requirements for mining; limited incentives for researchers to use the technique and a lack of understanding on the part of publishers. This last point was illustrated in an ALPSP report: “a large number of the publishers surveyed have little or no understanding of text mining, and many suggest in their comments that they have never been approached by a client about text mining” (Inger & Gardner, 2013). A 2016 study of the use of TDM for public research in the UK and France suggested relatively little had changed