

Opinion on Legality of JNU Data Depot of SCIS

Zakir Thomas, PhD

Background

I, Zakir Thomas, upon the request of Prof (Dr) Andrew Lynn, Jawaharlal Nehru University (JNU) have reviewed the Initial Concept Memorandum, Terms of Use, Memorandum of Association and Cooperation, and Tiered Architecture Description of the JNU Data Depot. I also had detailed discussion with Dr Andrew Lynn about the JNU Data Depot as I know Professor Lynn as he was one of the lead Project Investigators of Open Source Drug Discovery (OSDD) of the Council of Scientific and Industrial Research (CSIR), of which I was the Project Director.

This opinion I am furnishing is in my personal capacity and does not reflect the position of my employer, the Government of India or the Income Tax Department where I serve as a Commissioner of Income Tax.

Relevant Experience to Tender this Opinion

I have 30 years of experience working as an Indian Revenue Service officer of the Government of India in various capacities including Deputy Secretary in the Department of Higher Education in the Ministry of Human Resources Development where I was Registrar of Copyright for the Government of India. As Project Director of OSDD and Head of Director General's Technical Cell in CSIR for five years, I worked closely with senior scientific administrators of India and had the first-hand experience of the limitations faced by Indian researchers on access to knowledge

I possess an M Sc in Physics from Mahatma Gandhi University, Kerala, LLB from Delhi University, a Masters degree in Intellectual Property from the Franklin Pierce Law Centre of the University of New Hampshire, US. I have earned my Doctor of Philosophy (PhD) from the Indira Gandhi National Open University, with special focus on innovation in neglected diseases.

I have several peer reviewed publications and book chapters in reputed journals and books. I have represented India in the Standing Committee on Copyright and Related Rights of the World Intellectual Property Organisation (WIPO) and been invited speaker in a large number of academic conferences. I had been the Plenary Speaker in the World IP Congress 2016.

It may be noted that I am not a lawyer, and the opinion I am furnishing should not be taken as legal advice. This opinion is based on my analysis of the Indian copyright law based on my experience as the Registrar of Copyright, further readings and understanding of the subject.

Factual Matrix

I have perused the Agreement between School of Computational and Integrative Sciences (SCIS), Jawaharlal Nehru University and Public Resource.org, Inc. SCIS is established to innovate and integrate computational and analytical approaches adopted in different branches of sciences. SCIS has established a High-Performance Computing Facility (HPCF) towards this objective.

Access to HPCF is controlled and available within its confines to its researchers and remote access is available only to carefully chosen, select few researchers and faculty. Thus, this is purely an educational project of SCIS operating within the University department. The project has carefully established cybersecurity measures and terms of use to ensure that any persons accessing the JNU Data Depot are engaged only in research or education.

It is stated in the Agreement that one of the key limitations faced by the researchers is access to data sets. Therefore, a number of data sets have been loaded on the HPCF with the strict access limitations described above. Many of them are listed in the Agreement. The Nature article refers to the fact that over the past year, researchers have built a store of text and images extracted from 73 Million journal articles. The sources are various, but the origin of articles other than that mentioned in the Agreement is not clear. It is imputed in the Nature article that these may contain copyright protected articles accessed without permission. The articles are not available for free download. No copies are being made for distribution or consumption of any materials under copyright. Articles cannot even be searched using their DOI (article identifier) as per the Nature article. It is therefore a facility meant only for text and data mining.

Opinion

Based on the above facts I have analysed if the JNU Data depot referred above violates the Copyright Act, 1957. In my considered opinion, the JNU facility described above does not violate provisions of the Act, for the reasons stated in the ensuing paragraphs.

The JNU Data Depot at SCIS is clearly and purely an educational project. There is no dispute on this fact. The researchers are accessing this facility only for educational purposes. Even while accessing the articles, they are prohibited from downloading the articles. The facility is only providing access to researchers for text and data mining (TDM). The Agreement notes that the biggest impediment for researchers in doing TDM is lack of access to data. Therefore, as stated in the Nature article, data has been compiled from various sources by the researchers. The copyright status of these articles is not ascertainable, as to access this data set for such purpose itself is not a permitted act as per the agreement of access to HCF and access is available only for TDM.

Therefore, for the purpose of this opinion, it is assumed that a good number of the articles that are in the database have found a place there without authorisation from the rightsholders. So in this background I am analysing the copyright issues surrounding the JNU Data Depot from the following perspectives:

1. Is storing the data set containing copyright protected articles, without permission of the right holder, in the HPCF for TDM purposes illegal?
2. Is accessing the articles in the HPCF for TDM purposes by researchers illegal?
3. In view of the above two questions and in the background, is the JNU Data Depot illegal?

I am assuming for the purpose of this analysis that the JNU data depot is entirely consisting of copyright protected material (that obviously is not the case, but for the purpose of analysis this extreme position is adopted).

It is an undisputed fact that JNU Data Depot is purely an educational and research project. It is accessed and used only for research purposes and specifically for TDM. This research activity is within the mandate of the University and SCIS. Even if the JNU Data Depot contains copyright infringing material stored without permission of the rightsholder, the access to the data set and use by researchers and faculty and the provision of this facility by the University is a permitted act within the meaning of Copyright Act, 1957 (the Act).

Chapter XI of the Act deals with "Infringement of Copyright" (the chapter heading). Section 51 states that Copyright in a work shall be deemed to be infringed (a) if any person, without a licence, (i) does anything which is the exclusive right of the author, (ii) permits for profit any place to be used for communication of the work to the public; (b) when a person makes for sale or hire or distribute for the purpose of trade or import into India any infringing copy of the work. It is clear that the JNU facility is not used for any profit-oriented activity. It is an educational and research activity that is carried out in a not for profit public university. Therefore, the only way the facility can infringe the copyright is through violation of section 51 (a) (1) which prevents anyone from doing anything which is the exclusive right of the author, which includes storage and communication to the public. It might be argued prima facie that JNU Data Depot violates this section, but that argument is invalid for the reasons detailed below.

Section 52 of the Act is titled, 'Certain Acts not to be Infringement of Copyright'. This section permits acts that are other wise infringement (illegal acts). Without this section 52 these acts are to be carried out only with a licence of the author (for it would have violated section 51 discussed above). The Section 52 lists out a number of acts that shall not constitute infringement even if accessed violating copyright. In plain terms, even if the acts are infringement of copyright under 51 (illegal), if the act falls within the various acts listed in 52, they are legal as they are permitted by law.

Section 52 (1) (a) deals with 'a fair dealing with any work for the purposes of private personal use including research'. The activity carried out by JNU Data Depot at SCIS is research falling within the meaning of Section 51 (1) (a). Therefore, the researchers accessing the works (stored without authorization of the rightsholder) is permitted under the Act, and therefore legal. It is also a fact that the TDM cannot be carried out without accessing such data. The question then is, is the work is kept at JNU without the licence of the right-holder if it

constitutes infringement. The entire *raison d'être* of section 52 is to permit access to works without permission. Therefore if the data set lodged in the data repository is solely for research purposes and accessed solely for that purpose, and if the research cannot be carried out without accessing the data set, then such acts fall within the ambit of Section 52 and become permitted or legal acts. The project has carefully established cybersecurity measures and terms of use to ensure that any persons accessing the JNU Data Depot are engaged only in research in the form of TDM. The activities carried out are in the nature of non-consumptive use. No copies are being made for distribution or consumption of any materials under copyright. It is the law that for research purposes a researcher keeps an infringing copy in his personal possession, it is not an illegal act as per section 52 of the Act. For example, he may photocopy a literary work for research purposes. If many researchers possess such copies and exchange it for purely research purposes, even that does not constitute infringement. If these researchers keep such copies in the class room or at any other part of the institution for their perusal for the purpose of research it will still not constitute infringement and therefore it will be a legal and permitted act. Instead of physical copies the JNU Data Depot is having digital copies. Therefore, the research activity accessing an infringing digital material is permitted under law as these acts would still be 'fair dealing' with a work contemplated under Section 52. Exercising one's rights of fair dealing with a work is equally important and legal as protecting one's copyright under the Act. If we deny the researchers the permissions granted to them under section 52, it will be contravening the scheme of the Act itself.

Even if the number of the works kept in the Data Depot is large, it would still be covered under Section 52 as the access is only in the nature of 'fair dealing' as it is only for research in the nature of text and data mining. Researchers are not making further copies and distributing it; they are not allowed to do it. Thus, their research activity does not in any way adversely impact the prospect of a sale which the rightsholder otherwise could have made. It is a purely non-consumptive use for research purposes in an educational institution. The US Courts have permitted Google to scan and store thousands of copyrighted books without permission and display snippets from these books as part of its search service. Google does not allow them to be downloaded or read in their entirety. The facts of JNU Data Depot are exactly the same. As in the case of Google, there is no permission from authors. It is stored and kept by JNU in its servers and in the case of Google in their servers. The researchers in JNU are accessing only snippets of data and texts during their TDM research. In the case of Google the whole world accesses the snippets. In JNU only select few researchers access the data set. Therefore, the JNU data depot can be held more legal than even the Google Books which have already been held as legal. Following this legal precedent, it can be concluded that by creating the dataset containing infringing material in its HPCF at SCIS, JNU has not violated the provisions of the Act and this activity is legal within its framework.

The JNU Data Depot does not adversely affect the commercial prospect of the rightsholders at all. No rightsholder is making available a similar large scale collection for data mining available, at a fee or otherwise. To my knowledge no such data depot exists in the world which contain works from Gandhi to Tagore to Nehru to large corpus of scientific articles. A researcher who is required to do TDM for the purpose of his research could not carry out his research in any other part of the world even if he is willing to pay any amount. It is possible only by bringing together such a corpus which the JNU Data Depot has done. It has created it only for the purpose of education which is its mandate and students are individually accessing it which is their legal right under Section 52. I will not be surprised if a student challenge the institution in a Court if his access for research is denied as he is entitled to do so under Section 52 of the Act.

There are other provisions under section 52 which underlines the legality of JNU Data Depot. Section 52 (1) (o) allows a library to make up to three copies of the work for the use of the library provided it is a non-commercial library. HPCF at JNU is a noncommercial facility. The Section enables making copies of protected works which otherwise would have been illegal. The JNU Data Depot is only making digital copies. To my understanding the Data Depot is not having multiple copies of the same work and is having only a single copy. What is permitted inside the precincts of a library which had a physical context in the analogue world, is permitted in a Data Depot which has the characteristics of a library in the digital world. It is only natural to treat the Data Depot as a library as it contains as much if not more information that contains in a library and anyone in charge of the same, who lays down the rules of access can be considered the librarian. Libraries are accessed by researchers and so is the Data Depot. The philosophical underpinnings of this Section 51 (1) (o) justifies the acts carried out in JNU Data Depot.

It is an essential principle of copyright law that there is no copyright on facts and data per se. If a fact or a data per se is part of a protected work, they do not get protection just because the whole work enjoys protection. Therefore, accessing the data in an otherwise protected work is not an infringement. Making a copy of the work in the Data Depot to enable researchers to access just the data cannot be held illegal. Even in the physical world, if a property can only be accessed by trespassing another's property, the law permits easement rights. Without accessing these articles, the unprotected data and facts cannot be accessed. It is the right of a researcher to access the unprotected work. It is well within the rights of a University to provide the access to such data and facts. Anyone who denies such access is an accomplice in enabling illegal profiteering by the persons who claims to be rights holder over unprotected work. In fact, anyone who denies access to an unprotected work under the copyright law is doing an illegal act.

Text and data mining is legal for reason of public policy such as the advancement of science and because the only use is the transformative use of extraction of facts, materials that are not under copyright. This is an allowable use under Indian copyright law.

Another provision of law which supports the legality of the JNU Data Depot is Section 52 (1) (i) which deals with educational exemption. Reproduction of any work by a teacher or a pupil during the course of instruction is not an infringement. This section was judicially interpreted in Delhi University photocopy Case (See The Chancellor, Masters & Ors vs Rameshwari Photocopy Services & Ors, High Court of Delhi at New Delhi on 9 December, 2016) The Court interpreted 'during the course of instruction' broadly. The research activity carried out in JNU is an activity that is during the course of instruction in the university. Therefore the JNU Data Depot making copies of protected work is as much legal, if not more, as a photocopy shop in Delhi University making copies and distributing to students at a price. The case of JNU is more strong as there is no distribution of copies to students. The student researchers are only accessing a work kept in JNU under strict access conditions to carry out TDM research. They are not even allowed to download or even to do a DOI search nor to make copies.

Conclusion

The JNU Data Depot of SCIS is legal and does not infringe the provisions of Copyright Act, 1957, in my view. The relevant questions are answered as below:

Is storing the data set containing copyright protected articles, without permission of the right holder, in the HPCF for TDM purposes illegal?

Answer – No.

Is accessing the articles in the HPCF for TDM purposes by researchers illegal?

Answer – No

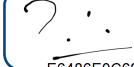
In view of the above two questions and in the background, is the JNU Data Depot illegal?

Answer – No

Based on the facts presented to me I have no hesitation in concluding that the activities carried out in the JNU HPCF Data Depot at SCIS are permitted acts and therefore legally valid acts.

Signed and Submitted in New Delhi:

DocuSigned by:



E6486E0C692C418...

Dr. Zakir Thomas

8/2/2019