LICENCES FOR EUROPE – WORKING GROUP 4 – TEXT AND DATA MINING

17 April 2013

Publishing Sector Proposal for a Sample Licence Clause for Text and Data Mining (TDM) of subscribed copyright-protected works and materials

Introduction and Context:

The sample clauses presented here are not a self-standing licence but to be inserted, added to or read with an existing electronic access and subscription agreement between a publisher and a university or other research institution (the "Licensee").

Subscription agreements provide researchers, employees, representatives of the licensee and students (defined as "Authorised Users") legal access to and user rights regarding certain subscribed copyright-protected works and other electronic materials, eg databases, (collectively referred to as the "subscribed Content"). Subscribed Content is usually identified in a schedule annexed to any subscription agreement. The clauses below would add text and data mining for non-commercial purposes to such subscription agreements.

Publishers and research organisations should consider agreeing to inserting these sample clauses or other equivalent clauses in existing agreements, or should consider otherwise to provide suitable text and data mining pilot agreements that enable non-commercial text and data mining, if only to gain experience valuable for researchers, technologists and publishers alike.

Where publisher include the sample clauses presented here (or equivalent clauses) or provide for pilot TDM agreements, they may also consider at the same time offering TDM "click-through API agreements for individual researchers who may have access under subscriptions or pilot agreements as so-called "authorised users".

SAMPLE CLAUSES

Clause 1 – Licensed uses: TDM

The Licensee may use all of the subscribed Content in the course and for the purpose of non-commercial research:

(a) for Text and Data Mining (TDM) by carrying out the following activities: download, extract and index information from the subscribed Content to which the Licensee has access under this Subscription Agreement. Where required, mount, load and integrate the results on a server used for the Licensee's text-mining system and evaluate and interpret the TDM Output for access and use by Authorised Users.

(b) store electronic copies of the subscribed Content as necessary solely to ensure efficient use by Authorized Users in connection with their work on a TDM project and only during the lifetime of any project, and provided locally stored copies are kept separate from proprietary or other third-party content and datasets. Such locally stored copies of subscribed Content are not to be redistributed for any purpose or to be used for marketing of the Licensee or the Authorised User's research capabilities. This also includes any commercial research capabilities.

(c) use TDM Output as part of original non-commercial research carried out by Authorised Users and describe or otherwise reproduce extracts and quotations from TDM Output as part of original works of authorship, eg research reports, research papers and research articles. Where subscribed Content is embodied, quoted or referred to, or where bibliographic metadata of subscribed Content is displayed, it should be accompanied by a DOI link that points back to the individual full text item of subscribed Content.

(d) make the results of any TDM Output available on an externally facing server or website beyond the inclusion of a few lines of query-dependent text of individual full text items of subscribed Content (eg extracts from articles or book chapters) which shall be in any event shorter than 150 characters or 15 words or 1 complete sentence or limited to bibliographic metadata and in no event shall the TDM Output contain links to access substantial parts of a full-text work or database of subscribed Content beyond the above limitation;

(e) Subject to sub-clause 1(d) above and clause 2 below, the Licensee may for non-commercial purposes communicate the TDM Output whether for internal or external use without creating products that substitute for subscribed Content, provided that the Licensee may not make the results or the TDM Output in the form of data sets available to third parties for their use.

Clause 2 – Prohibited uses relating to subscribed Content and TDM Output

[*Explanatory Note: The grant of rights regarding permitted uses of TDM contained in Clause 1 above in no way broadens the scope of other rights granted under the subscription agreement and the grant of rights remains subject to each of the limitations and exclusions, including regarding third-party access to subscribed Content. To any subscription agreement may be added the following*:]

(a)	Subject to clause 1 above and unless agreed by separate written agreement, the Licensee and its Authorized Users may not engage in activities listed under (b.1 to b.8).

(b.1)	create derivative products or services that would compete with or be substitutable for the Publishers' products and services;

(b.2)	extract, develop or use subscribed Content in any commercial activity;

(b.3)	abridge, modify, translate subscribed content, or create any derivative work based on the subscribed Content, except to the extent necessary to make them perceptible on a computer screen to Authorized Users;

(b.4)	allow a third party to harvest any TDM Output to an internal server;

 (b.5)	remove, obscure or modify in any way any copyright notices, other notices or disclaimers as they appear in the subscribe Content;

(b.6)   In the case of Text and Data Mining at Publisher's site: use any robots, spiders, crawlers or other automated downloading programs, algorithms or devices to continuously and automatically search, scrape, extract, deep link, index or disrupt the working of the subscribe Content;

(b.7)   substantially or systematically reproduce, retain or redistribute the subscribed Content.

c) This clause 2, as well as the restrictions in sub-clause 1 (c) to 1(e) shall remain effective and survive termination of this Subscription Agreement, howsoever arising.

Clause 3 – Security, Grant of Access Rights, Formats and Delivery Mechanisms

The Publisher and the Licensee agree the following access protocols:

(a) The Licensee shall ensure compliance with Publisher's Usage policies, including security and technical access requirements. Text and data mining may be undertaken on either locally loaded subscribed Content or as mutually agreed.

[*INSERT AS APPLICABLE: onsite delivery, ftp file transfer, format (eg XML), application programming interfaces (APIs), use of "free" Publisher Content at Publisher website, including article abstracts*].

(b) Licensee shall implement and maintain adequate and effective state of the art data security systems and measures, in line with international industry standards and best practice.

(c) In the event the Licensee collects or otherwise processes or uses personal data, Licensee shall expressly obtain consent from the end user or any person that is a beneficiary of applicable data protection laws which shall be limited to the use the end user/affected person expressly consents to. Licensee shall undertake the appropriate technical and security measures to protect any personal data Licensee collects.

(d) The Publisher may require the Licensee and Authorised Users to adhere to scheduling and staggering over time any TDM activities in line with publishers' policies referred to in clause 1.1.2 above.

(e) The Publisher does not provide any warranty or assurance of fitness or availability of Publisher'' electronic servers and systems to carry out TDM or to extract TDM Output, nor of completeness or accuracy of any downloaded materials.

Definitions

"Text and Data Mining" means to perform extensive automated searches of subscribed Content, including data embodied therein, the sorting, parsing, addition or removal of linguistic structures, and the selection and inclusion of subscribed content into an index or database for purposes of classification or recognition of relations and associations."

"TDM Output" means the result of any Text and Data Mining activity or operation, capable of fixation, reproduction and/or communication in any form, including without limitation the creation of an index, reference, abstract, relative or absolute description or representation of subscribed Content, an algorithm, formula, metrics, method, standard or taxonomy describing or based on subscribed Content, a relational expression or measurement, whether scalable or not, of subscribed Content, extraction, alternative representation or translation, expression or discussion of any

extracts from mined subscribed Content, whether in the form of a direct extraction or a representation in any form which is based on subscribed Content.

<div align="center">***</div>

Additional Definitions

[unless already covered in any subscription agreement to which above clauses relate]

"Authorised User" means researcher, employee or representative of the Licensee and any student enrolled at the Licensee.

"Licensee" is a [public university or institution of higher-learning or a publicly funded non-profit research institution or a public research library].

"Subscribed Content" means the whole or any part of a copyright-protected literary work, of a database or of any other information product owned or controlled by the Publisher or its suppliers, and includes (without limitation) any illustration, table, figure, header information, abstracts, bibliographic information or any video or computer-animated design embedded or incorporated in such literary work, database or information product.

<div align="center">***</div>