



Il diritto di *Text and Data Mining* (TDM) non esiste

Salvatore Orlando

L'a. analizza le disposizioni della direttiva copyright nel mercato unico digitale (direttiva (UE) 2019/790: CDSMD) relative alla c.d. eccezione di *Text and Data Mining* al fine di stabilire se per consentire le attività tecniche implicate dalle operazioni di "estrazione di testi e di dati" (TDM) ci sia sempre e necessariamente bisogno o di una eccezione o di una licenza come contemplato dagli artt. 3 e 4 CDSMD. Sulla base di una interpretazione in linea con la più recente giurisprudenza della Corte di Giustizia dell'Unione europea sulla nozione di riproduzione (in particolare il caso *Pelham*), l'a. argomenta per la conclusione negativa, affermando che la riproduzione di dati leggibili solo dagli elaboratori (*machine readable data*) non comporta un'attività di riproduzione o di estrazione riservata dai diritti menzionati negli artt. 3 e 4 CDSMD. L'a. osserva dunque che le attività di analisi automatizzata di testi e di dati in formato digitale (ossia le attività di TDM, secondo la relativa definizione), avendo ad oggetto normalmente dati leggibili solo dagli elaboratori non comportano nemmeno ordinariamente un problema di interferenza con i diritti considerati dagli artt. 3 e 4 della direttiva (UE) 2019/790. L'a. ulteriormente argomenta che nel diritto della UE i "dati" non formano oggetto dei diritti di riproduzione ed estrazione menzionati nella direttiva in commento né di altri diritti di proprietà intellettuale. Su queste basi, l'a. conclude affermando l'inesistenza di un diritto esclusivo di TDM.

Direttiva sul diritto d'autore nel mercato unico digitale (CDSMD) – Estrazione di testo e dati (TDM)
– Intelligenza artificiale – Diritto di riproduzione – Diritto *sui generis* – Diritto di estrazione

SOMMARIO: 1. Premessa – 2. L'annuncio del Governo del Regno Unito di una prossima eliminazione di qualunque restrizione alla c.d. eccezione di *Text and Data Mining* (TDM) – 3. Lo studio del 2022 della Commissione europea su copyright e nuove tecnologie – 4. L'interpretazione della direttiva copyright nel mercato unico digitale (CDSMD) ed il rischio di creare regole (inesistenti) dalle eccezioni – 5. Nella legislazione europea, i "dati" non formano oggetto dei diritti esclusivi menzionati dagli artt. 3 e 4 CDSMD né di altri diritti di proprietà intellettuale – 6. Il rischio di equivoco linguistico sulla parola "estrazione": l'esclusiva rilevanza (ai fini della c.d. eccezione di TDM) del diritto di estrazione previsto dalla direttiva banche dati 96/9/CE e la diversa nozione di "mining" nella CDSMD – 7. Relativamente alla sola disciplina italiana di recepimento della CDSMD (art. 70-ter, co. 1, l.a.): l'inesistenza di un generale divieto di comunicazione al pubblico dei risultati delle ricerche compiute in esito all'applicazione di tecniche di TDM – 8. Il sillogismo alla base della c.d. eccezione di TDM – 9. Non è vero che il TDM comporti necessariamente attività di riproduzione o estrazione riservate ex lege nella legislazione europea – 10. Il concetto di riproduzione e i dati leggibili solo dagli elaboratori elettronici – 11. Non è vero che il TDM comporti nemmeno ordinariamente attività di riproduzione o estrazione riservate ex lege nella legislazione europea – 12. Le tendenze del nuovo diritto europeo sui

S. Orlando è professore ordinario di Diritto privato presso Sapienza - Università di Roma e direttore dell'OGID (Osservatorio Giuridico sull'Innovazione Digitale).

Lo scritto riprende, con integrazioni e note, la relazione dal titolo *L'eccezione di Text and Data Mining e le nozioni legali di "informazioni" e "dati" nella legislazione del mercato unico digitale* svolta dall'a. al Convegno "Ricerca, educazione e accesso al patrimonio culturale. Un confronto tra diritti fondamentali ed eccezioni al diritto d'autore", tenutosi a Roma presso il Consiglio Nazionale delle Ricerche (CNR), nell'Aula Marconi, il 6 maggio 2022.

Questo saggio fa parte della sezione monografica "Ricerca, educazione e accesso al patrimonio culturale. Un confronto fra diritti fondamentali ed eccezioni al diritto d'autore" a cura di Deborah De Angelis, Sebastiano Faro e Ginevra Peruginelli.



dati e la diversa interpretazione della CDSMD in funzione di protezione dei diritti di proprietà intellettuale – 13. Il TDM come risorsa (invece che pericolo) per l'industria del copyright – 14. Conclusioni

1. Premessa

Nel presente intervento analizzeremo le disposizioni della direttiva (UE) 2019/790, c.d. direttiva copyright nel mercato unico digitale (di seguito anche “direttiva CDSM” o solo “CDSMD” dal suo acronimo in lingua inglese)¹ relative alla c.d. eccezione di *Text and Data Mining* (artt. 3² e 4³ CDSMD), al fine di stabilire se per consentire le attività tecniche poste in essere ai fini delle operazioni di *Text and Data Mining* (TDM) ci sia sempre e necessariamente bisogno o di una eccezione di legge o di una licenza nelle forme di legge⁴.

La conclusione che ne trarremo, per i motivi di seguito illustrati, è negativa. Tale conclusione può generalmente tradursi nell’affermazione dell’inesistenza di un diritto esclusivo di TDM.

2. L’annuncio del Governo del Regno Unito di una prossima eliminazione di qualunque restrizione alla c.d. eccezione di *Text and Data Mining* (TDM)

Il 28 giugno 2022, l’*Intellectual Property Office* del Regno Unito ha pubblicato un documento in esito ad una consultazione pubblica avviata il 29 ottobre 2021 avente ad oggetto le seguenti tre questioni: 1) se e come modificare il regime del copyright in relazione ai contenuti generati dagli elaboratori elettronici (*computer-generated works*); 2) se e come modificare il regime del c.d. *Text and Data Mining*; 3) se e come modificare il regime delle invenzioni e dei brevetti in relazione agli output di sistemi di intelligenza artificiale⁵.

In esito alla consultazione pubblica, la posizione del Governo del Regno Unito, come dichiarata nel citato documento del 28 giugno 2022 (di seguito il “Documento del 28 giugno 2022”) è nel senso di non introdurre allo stato alcuna modifica alle normative del Regno Unito riguardanti le questioni *sub* 1) e 3), ma di innovare il regime del c.d. *Text and Data Mining* (questione *sub* 2)), nel senso di eliminare qualsiasi restrizione alle attività di TDM fondata sul diritto di autore e sul diritto *sui generis* sulle banche dati attraverso l’introduzione di una nuova eccezione a tali diritti che consenta le attività di TDM per qualsiasi finalità. Nel Documento del 28 giugno 2022, viene annunciato che il Governo del Regno Unito individuerà le modifiche legislative più adeguate al fine di conseguire questo obiettivo senza ritardo.

Per comprendere l’importanza della posizione annunciata dal Governo del Regno Unito nel Documento

del 28 giugno 2022, si deve ricordare che il Regno Unito aveva già emanato una specifica normativa prima della direttiva CDSM⁶. Si tratta delle norme (ancora in vigore in attesa delle modifiche annunciate dal Documento del 28 giugno 2022) contenute nella sezione 29A del *Copyright, Designs and Patents Act 1988* rubricata “Copie per analisi di testo e di dati per ricerca non commerciale” (*Copies for text and data analysis for non-commercial research*) che, con una serie di specificazioni e condizioni, seguono l’impostazione per la quale l’eccezione si applica solo limitatamente alle copie effettuate per finalità di ricerca non commerciale⁷.

Le attività di *Text and Data Mining* sono definite nel Documento del 28 giugno 2022 come (nostra traduzione) «l’uso di tecniche computazionali per analizzare grandi quantità di informazioni al fine di individuare modelli, tendenze ed altre informazioni utili»⁸. Nella direttiva CDSM la definizione è la seguente: «“estrazione di testo e di dati” (text and data mining): qualsiasi tecnica di analisi automatizzata volta ad analizzare testi e dati in formato digitale avente lo scopo di generare informazioni inclusi, a titolo non esaustivo, modelli, tendenze e correlazioni» (art. 2 CDSMD).

Degno di nota è che il Documento del 28 giugno 2022 abbia posto al centro della questione l’importanza delle attività di TDM per lo sviluppo dei sistemi di intelligenza artificiale, come si vede dallo stesso *wording* del quesito oggetto della consultazione pubblica: *Licenze o eccezioni per il TDM, che è spesso rilevante per l’uso e lo sviluppo dell’IA* (e v. i punti da 31 a 62 del Documento del 28 giugno 2022), e, ancor più significativamente nel punto 62 del Documento del 28 giugno 2022: «The Government’s ambition is to make the UK a global centre for AI innovation. The new exception will ensure the UK’s copyright laws are among the most innovation-friendly in the world [...]».

Le opzioni regolamentari che erano state sottoposte alla consultazione pubblica erano le seguenti: opzione 0 = nessun cambiamento rispetto all’assetto normativo esistente, ovvero mantenere l’attuale eccezione limitata alle copie per ricerca non commerciale; opzione 1 = modificare le regole sulle licenze relativamente al TDM; opzione 2 = estendere l’eccezione alla ricerca commerciale; opzione 3 = estendere l’eccezione a qualsiasi scopo, con facoltà di *opt-out* in favore dei titolari dei diritti; opzione 4 = estendere l’eccezione a qualsiasi scopo, senza facoltà di *opt-out* in favore dei titolari dei diritti.



In esito alla consultazione pubblica, la posizione del Governo del Regno Unito è stata nel senso dell'opzione 4, ed è stata motivata come segue (nostra traduzione): «(59) L'introduzione di una eccezione che si applica al TDM commerciale porterà benefici a un'ampia platea di stakeholders nel Regno Unito, tra cui ricercatori, sviluppatori di IA, piccole imprese, istituzioni di tutela del patrimonio culturale, giornalisti e cittadini impegnati in attività civicamente rilevanti [engaged citizens]. Prodotti e servizi disegnati per i clienti [targeted products and services] governeranno alle imprese e ai clienti. I risultati della ricerca potranno giovare anche al più ampio pubblico. Ciò potrebbe accadere, ad esempio, supportando la ricerca e l'innovazione nella salute pubblica. Alcuni utilizzano il TDM e l'IA anche nei settori industriali legati alla creatività per comprendere il mercato o creare nuove opere – anche essi vedranno benefici. I benefici ridurranno il tempo necessario per ottenere l'autorizzazione da molteplici titolari di diritti e non saranno dovute commissioni di licenza. Ciò comporterà un'accelerazione del TDM e dello sviluppo della IA. (60) Questi cambiamenti valorizzano al meglio le possibilità conseguenti alla Brexit. Esse aiuteranno a rendere il Regno Unito più competitivo come sede di stabilimento per aziende che fanno data mining. (61) I titolari di diritti non potranno più chiedere compensi per licenze rette dalla legge del Regno Unito a titolo di TDM e non potranno negoziare o esercitare facoltà di *opt-out* per l'eccezione. Il nuovo regime può anche avere conseguenze per coloro che hanno costruito modelli di impresa anche intorno alle licenze di dati. Tuttavia, i titolari di diritti manterranno salvaguardie per proteggere i loro contenuti. La maggiore salvaguardia consisterà nel requisito di un accesso legittimo. Ciò sta a significare che i titolari dei diritti possono scegliere la piattaforma dalla quale essi rendono le loro opere accessibili, e possono chiedere compensi per l'accesso attraverso abbonamento o per singoli accessi. Essi potranno anche adottare misure per assicurare l'integrità e la sicurezza dei loro sistemi. (62) L'ambizione del Governo è di fare del Regno Unito un centro mondiale per l'innovazione dell'IA [...]».⁹

3. Lo studio del 2022 della Commissione europea su copyright e nuove tecnologie

Per quanto riguarda il diritto euro-unitario, come risultante dal sistema della c.d. eccezione di TDM designata dalla CDSMD giova segnalare sin da subito un recente studio commissionato dalla Commissione

europea, intitolato *Study on copyright and new technologies: copyright data management and artificial intelligence*¹⁰, nel quale si trovano alcune interessanti osservazioni dedicate al TDM.

Da questo studio emerge innanzitutto la piena consapevolezza dell'importanza delle attività automatizzate di analisi dei dati per lo sviluppo dei sistemi di intelligenza artificiale¹¹.

Tale consapevolezza non è viceversa manifestata nella CDSMD, nella quale nessun articolo e nessun considerando menzionano l'intelligenza artificiale (né a proposito dell'eccezione di TDM né a proposito di altri istituti)¹².

In secondo luogo, in questo studio si mette in dubbio che qualsiasi applicazione di TDM comporti di per sé una violazione dei diritti di proprietà intellettuale interessati dal sistema delle eccezioni e delle limitazioni introdotto dagli artt. 3 e 4 CDSMD¹³. In particolare, in questo studio si espone il dubbio che le attività strumentali al TDM siano tali da violare sempre il diritto di riproduzione, come contemplato dalla legislazione euro-unitaria rilevante ai sensi degli artt. 3 e 4 CDSMD. A tal proposito, si afferma essere rilevante stabilire se l'opera protetta sia riconoscibile (*recognisable*) e dunque sfruttata (*exploited*) nel corso del processo di TDM o successivamente, nel senso che solo in questo caso possa predicarsi un'interferenza potenzialmente lesiva del diritto di riproduzione in assenza di elementi di fatto integranti un'eccezione o una licenza ai sensi degli artt. 3 e 4 CDSMD¹⁴.

Nel medesimo studio, contemporaneamente, si prende posizione in favore di quella che viene definita come esigenza di rivisitazione del concetto di riproduzione, dichiaratamente in linea con parte della dottrina¹⁵, e si dà conto della sentenza della Corte di Giustizia nel caso *Pelham*¹⁶, come precedente citato a supporto di una simile rivisitazione nel segno del criterio della riconoscibilità¹⁷.

Il dubbio sollevato in questo studio appare meritevole di approfondimento sia per le sue ricadute pratiche (relativamente allo sviluppo della ricerca e dell'impresa nel campo dell'intelligenza artificiale) sia per le implicazioni sistematiche (relativamente alla tassonomia, alla natura e al contenuto dei diritti di proprietà intellettuale) sia per l'incisione delle libertà e dei diritti fondamentali interessati dai temi della conoscenza nell'era digitale (relativamente alle effettive possibilità di formazione, fruizione e trasmissione della conoscenza nell'era dei dati digitali, implicate dall'accesso alle tecniche di analisi automatizzata di dati digitali e dalla circolazione dei relativi risultati).



4. L'interpretazione della direttiva copyright nel mercato unico digitale (CDSMD) ed il rischio di creare regole (inesistenti) dalle eccezioni

A proposito del sistema della c.d. eccezione di TDM nella CDSMD (artt. 3 e 4 CDSMD)¹⁸, si deve evidenziare innanzitutto, dal punto di vista ermeneutico, il rischio di creare regole dalle eccezioni, ossia il rischio che non interpretando correttamente la nuova disciplina se ne facciano derivare norme inesistenti.

Una lettura superficiale della CDSMD può portare alla conclusione che le attività di TDM non possano essere legittimamente poste in essere se non nei ristretti limiti soggettivi ed oggettivi contemplati dagli artt. 3 e 4 della medesima direttiva. Questa è una interpretazione sbagliata della nuova disciplina, la quale, invece, segue lo schema dell'eccezione, tale per cui, alcune attività riservate ai titolari di certi diritti esclusivi (che hanno ad oggetto tali attività) possono eccezionalmente essere poste in essere da terzi nella cornice di determinate circostanze. La nuova disciplina in commento consente eccezionalmente a determinati terzi di porre in essere alcune attività riservate ai titolari di certi diritti esclusivi se ricorrono determinate circostanze, tra le quali la circostanza che le medesime attività siano *poste in essere "ai fini" di TDM ("ai fini dell'estrazione di testo e di dati")*. Dunque il TDM (ossia l'estrazione di testo o di dati, ossia l'analisi automatizzata di dati in formato digitale) non è – nello schema logico della nuova disciplina – l'attività riservata ai titolari di diritti esclusivi, bensì è un elemento di fatto che concorre insieme ad altri a consentire eccezionalmente a terzi di porre in essere certe attività riservate ai titolari di diritti esclusivi. Laddove le attività riservate vengano poste in essere ai fini del TDM, allora, ricorrendo certe ulteriori condizioni, devono avere applicazione i regimi dell'eccezione dell'art. 3 e delle eccezioni e/o le limitazioni dell'art. 4 CDSMD.

L'errore sopra denunciato è tuttavia ricorrente, con la conseguenza che non di rado si trova affermato che il TDM (ossia l'analisi automatizzata di dati in formato digitale) sarebbe l'attività riservata che può essere eccezionalmente esercitata da terzi nelle ristrette circostanze di cui agli artt. 3 e 4 CDSMD¹⁹. Da cui anche una serie di commenti allarmati, che, effettivamente, sarebbero giustificati se questa interpretazione della nuova disciplina fosse corretta²⁰; mentre l'allarme va rivolto verso la pervasiva ed inquietante diffusione dell'interpretazione errata, perché è ormai ricorrente sentir parlare di "diritti di TDM" nel contesto di "licenze di TDM"²¹. Nulla di più sbagliato.

Tutto al contrario, bisogna riaffermare che le attività riservate sono invece soltanto la "estrazione" e la "riproduzione" e non le attività di TDM. E qui c'è un altro pericolo di confusione, perché alle parole "estrazione" e "riproduzione" non può attribuirsi qualsiasi significato, anche eventualmente tratto dal linguaggio comune. Si tratta di attività che hanno una precisa definizione giuridica in quanto formano precisamente gli *oggetti* dei diritti esclusivi omonimi assoggettati alle eccezioni e limitazioni della nuova disciplina: il "diritto di estrazione" e il "diritto di riproduzione".

Ed infatti, e specificamente, nel regime dell'art. 3 CDSMD si tratta:

- dell'eccezione al diritto di riproduzione come previsto dall'articolo 5, lettera a) della direttiva banche dati 96/9/CE;
- dell'eccezione al diritto di estrazione come previsto dall'articolo 7, paragrafo 1, della direttiva banche dati 96/9/CE;
- dell'eccezione al diritto di riproduzione come previsto dall'articolo 2 della direttiva InfoSoc 2001/29/CE;
- dell'eccezione al diritto di riproduzione come previsto dall'articolo 15, paragrafo 1, della medesima CDSMD (pubblicazioni giornalistiche, utilizzo online);

Nel regime dell'art. 4 CDSMD, si tratta:

- delle eccezioni o limitazioni al diritto di riproduzione come previsto dall'articolo 5, lettera a) della direttiva banche dati 96/9/CE;
- delle eccezioni o limitazioni al diritto di estrazione come previsto dall'articolo 7, paragrafo 1, della direttiva banche dati 96/9/CE;
- delle eccezioni o limitazioni al diritto di riproduzione come previsto dall'articolo 2 della direttiva InfoSoc 2001/29/CE;
- delle eccezioni o limitazioni al diritto di riproduzione come previsto dall'articolo 4, paragrafo 1, lettere a) e b), della direttiva software 2009/24/CE;
- delle eccezioni o limitazioni al diritto di riproduzione come previsto dall'articolo 15, paragrafo 1, della medesima CDSMD (pubblicazioni giornalistiche, utilizzo online).

Dunque, quando gli artt. 3 e 4 CDSMD prevedono che gli Stati membri debbano disporre, rispettivamente, un regime di eccezioni (art. 3) e di eccezioni e limitazioni (art. 4) ai diritti esclusivi sopra elencati per le "riproduzioni" e le "estrazioni" effettuate da terzi "per fini di estrazione di testo e dati" da opere e altri materiali protetti, intendono dire che le attività di "riproduzione" e le attività di "estrazione" in questione sono esattamente quelle attività riservate ai titolari degli stessi diritti esclusivi ivi menzionati,



ossia sono precisamente *ed esclusivamente* le attività che formano l'oggetto di quei diritti esclusivi, con la conseguenza che se i terzi pongono in essere attività diverse da quelle di riproduzione ed estrazione riservate, la nuova disciplina non ha campo, perché non si tratta di attività riservate. Per comprenderci: se terzi pongono in essere attività di TDM esterne ad attività di riproduzione ed estrazione in senso tecnico, la disciplina non ha campo, perché l'attività di TDM non è di per sé riservata.

Pertanto, le attività di TDM, lungi dall'essere l'attività riservata su cui la nuova disciplina disegna eccezioni e limitazioni, è un'attività che può consentire un'eccezione o una limitazione dei diritti esclusivi che hanno ad oggetto le (diverse) attività che ne formano l'oggetto, e che sono le uniche attività riservate, ossia la riproduzione e l'estrazione in senso tecnico.

Per comprendere rettamente la nuova disciplina, è, pertanto, assolutamente (ed esclusivamente) necessario rivolgersi a quei diritti esclusivi per lumeggiarne, innanzitutto, l'oggetto.

Nei successivi paragrafi, evidenzieremo che le attività di riproduzione e di estrazione che formano l'oggetto degli omonimi e sopra elencati diritti esclusivi non hanno a loro volta ad oggetto i "dati", e, proseguendo, proveremo a chiarire ulteriori equivoci ai fini della esatta intelligenza del contenuto dei medesimi diritti esclusivi, e, con ciò, del significato da attribuirsi alla nuova disciplina.

5. Nella legislazione europea, i "dati" non formano oggetto dei diritti esclusivi menzionati dagli artt. 3 e 4 CDSMD né di altri diritti di proprietà intellettuale

Si deve evidenziare con nettezza che i dati non formano oggetto dei diritti esclusivi menzionati negli artt. 3 e 4 CDSMD²², né formano l'oggetto di qualsiasi altro diritto di proprietà intellettuale ai sensi della legislazione euro-unitaria²³.

Cominciamo con la direttiva banche dati 96/9/CE, ed esaminiamo separatamente il diritto di riproduzione e quello di estrazione ivi previsti.

Il diritto di riproduzione previsto dall'articolo 5, lettera a) della direttiva banche dati 96/9/CE ha ad oggetto infatti la "forma espressiva" della banca dati tutelabile mediante diritto d'autore; e la banca dati è definita ai sensi di quella direttiva come "raccolta" di determinate elementi – da non confondersi quindi con i medesimi elementi – in particolare come "raccolta di opere, dati ed altri elementi indipendenti"; ed infine tale raccolta può qualificarsi come banca dati se gli

elementi in essa raccolti sono «sistematicamente o metodicamente disposti ed individualmente accessibili grazie a mezzi elettronici o in altro modo» (art. 1, par. 2 direttiva banche dati 96/9/CE).

Quindi, qualsiasi cosa voglia dire "dati" ai sensi di questa disciplina (come noto, la direttiva banche dati 96/9/CE non contiene una definizione di "dati"), la banca dati, ai sensi della medesima disciplina, non si identifica con i dati, in quanto la "banca dati" è una "raccolta" di elementi, tra cui, eventualmente, i "dati", e deve necessariamente avere, per di più, per essere tale, determinate caratteristiche, come sopra ricordato.

Ed infine, come pure già ricordato, è testuale che il "diritto di riproduzione" (che riguarda solo la banca dati tutelabile con il diritto d'autore) ha ad oggetto la "forma espressiva" della banca dati tutelabile con il diritto d'autore, ossia la forma espressiva di una simile raccolta.

La definizione di banca dati appena richiamata è sufficiente anche per escludere che i dati siano oggetto del "diritto di estrazione" previsto dal successivo art. 7, paragrafo 1, della direttiva banche dati 96/9/CE. L'art. 7 della direttiva banche dati 96/9/CE prevede il diritto *sui generis* sulle banche dati, ossia il diritto per il costituente della banca dati di vietare a terzi «operazioni di estrazione e/o reimpiego della totalità o di una parte sostanziale del contenuto della stessa, valutata in termini qualitativi o quantitativi, qualora il conseguimento, la verifica o la presentazione di tale contenuto attestino un investimento rilevante sotto il profilo qualitativo o quantitativo». Ai sensi dell'art. 7, paragrafo 2, lett. a) della direttiva banche dati 96/9/CE «per "estrazione" si intende il trasferimento permanente o temporaneo della totalità o di una parte sostanziale del contenuto di una banca di dati su un altro supporto con qualsiasi mezzo o in qualsivoglia forma».

Più avanti evidenzieremo che questa nozione di estrazione non deve essere confusa con la nozione di TDM anche se nella lingua italiana la parola "mining" viene tradotta in entrambi i luoghi con "estrazione": nel primo caso per estrazione (*extraction* in inglese) si intende un'operazione di trasferimento della totalità o di una parte sostanziale del contenuto di una banca dati su altro supporto, nel secondo caso (*mining* in inglese) per estrazione si intende analisi automatizzata. Ai fini di questo paragrafo è sufficiente evidenziare che nemmeno questo diritto esclusivo ha ad oggetto i dati.

Con ciò abbiamo dimostrato che il diritto di riproduzione e il diritto di estrazione previsti dalla direttiva banche dati 96/9/CE richiamati agli artt. 3 e 4 CDSMD non hanno ad oggetto dati.



Nemmeno il diritto di riproduzione previsto dall'art. 2 della direttiva InfoSoc 2001/29/CE (richiamato dagli artt. 3 e 4 CDSMD) ha ad oggetto dati, in quanto esso ha ad oggetto, testualmente: le “opere”, relativamente al diritto di riproduzione attribuito agli autori delle opere (lettera a); le “fissazioni delle loro prestazioni artistiche”, relativamente al diritto di riproduzione attribuito agli artisti interpreti o esecutori delle prestazioni artistiche (lettera b); le “riproduzioni fonografiche” relativamente al diritto di riproduzione attribuito ai produttori di fonogrammi (lettera c); “l'originale e le copie delle loro pellicole”, relativamente al diritto di riproduzione attribuito ai produttori delle prime fissazioni delle pellicole (lettera d); le “fissazioni delle loro trasmissioni” relativamente al diritto di riproduzione attribuito agli organismi di radiodiffusione televisiva (lettera e).

Neanche il diritto di riproduzione previsto dall'art. 4 par. 1 lett. a) e b) della direttiva software 2009/24/CE ha ad oggetto “dati”, poiché, in entrambe le previsioni, è fatto chiaro che oggetto della riproduzione è un “programma per elaboratore”.

Infine, non hanno ad oggetto “dati” bensì hanno ad oggetto le “pubblicazioni di carattere giornalistico” i diritti ai quali si riferisce l'art. 15, paragrafo 1 della medesima CDSMD (richiamato anch'esso dagli artt. 3 e 4 della CDSMD).

Sembra opportuno aggiungere che diritti di proprietà intellettuale aventi ad oggetto dati non sono contemplati da nessuna altra norma del diritto dell'Unione. Nessuna norma dell'erigendo diritto euro-unitario dei dati – con questa espressione ci riferiamo alle molte discipline di recente conio, quali già emanate quali ancora in cantiere, che hanno ad oggetto i rapporti giuridici nell'ecosistema digitale²⁴ – prevede diritti esclusivi aventi ad oggetto dati.

Ai fini del nostro discorso sembra rilevante e sufficiente citare la Direttiva *Open Data* (UE) 2019/1024²⁵, il Regolamento *Data Governance Act* (UE) 2022/868²⁶, la proposta di regolamento *Data Act*²⁷ e la proposta di regolamento *Artificial Intelligence Act*²⁸.

In particolare:

- la Direttiva *Open Data* (UE) 2019/1024, esclude dal suo campo di applicazione, *inter alia*, i «documenti su cui terzi detengono diritti di proprietà intellettuale» (art. 1)²⁹,
- il Regolamento *Data Governance Act* (UE) 2022/868 contiene una definizione di dati (art. 2, par. 1 n. 1)³⁰ e alcune previsioni sui dati detenuti dal settore pubblico relativi a beni immateriali che formano oggetto di diritti di proprietà intellettuale (considerando 6, 10, 12, 18, 20³¹, artt. 2³², 3³³, 5³⁴, 7³⁵) e, per quanto riguarda la disciplina

dei “servizio di intermediazione dei dati”, esclude dalla relativa nozione i servizi di intermediazione di “contenuti” protetti dal diritto di autore (*copyright-protected content*: art. 2, par. 1, n. 11³⁶);

- la proposta di regolamento *Data Act* contiene la stessa definizione di dati di cui al citato regolamento *Data Governance Act* (UE) 2022/868³⁷, una norma (art. 35³⁸) intesa espressamente ad escludere la protezione del diritto *sui generis* previsto dall'art. 7 della direttiva banche dati 96/9/CE per i dati generati dall'uso dei prodotti (come ivi definiti)³⁹ e dei servizi correlati (come ivi definiti)⁴⁰, e, per il resto, prevede un generale riferimento alle istanze di tutela dei diritti di proprietà intellettuale (v. considerando 17⁴¹, 28⁴², 77⁴³), ove eventualmente rilevanti nel contesto dei vari settori di intervento coperti da questo strumento normativo⁴⁴;
- la proposta di regolamento *Artificial Intelligence Act* si occupa (*inter alia*) dei requisiti dei *data set*, tra cui i requisiti di completezza (art. 10), che possono essere indirettamente rilevanti nell'ambito di un'analisi generale sulle politiche e le finalità da perseguirsi nell'ambito della “Strategia europea dei dati” come dichiarate delle istituzioni dell'UE, da condursi in un'ottica di coerenza interna, e dunque anche nel senso della promozione di interpretazioni funzionalmente e sistematicamente orientate.

6. Il rischio di equivoco linguistico sulla parola “estrazione”: l'esclusiva rilevanza (ai fini della c.d. eccezione di TDM) del diritto di estrazione previsto dalla direttiva banche dati 96/9/CE e la diversa nozione di “mining” nella CDSMD

Si è già osservato che relativamente ai diritti esclusivi interessati dal sistema delle eccezioni e limitazioni a cui si riferiscono gli artt. 3 e 4 CDSMD, la parola “estrazione” (*extraction* in inglese) è utilizzata nel diritto euro-unitario soltanto ed esclusivamente nella direttiva sulle banche dati del 1996 (direttiva 96/9/CE) per indicare un'attività, ivi definita *ad hoc*, che concorre a formare insieme a quella di “reimpiego”, anch'essa ivi definita, l'oggetto del diritto *sui generis* sulle banche dati previsto dalla medesima direttiva.

In tale fonte (che, come pure si è già detto, è l'unica normativa rilevante per identificare il significato della parola “estrazione” in quanto designante il contenuto dell'unico diritto omonimo interessato dalla c.d. eccezione TDM) per “estrazione” si intende il «trasferimento [...] su un altro supporto» della «totalità o di una parte sostanziale del contenuto»



di una «raccolta di opere, dati o altri elementi», avente determinate caratteristiche, e rientrante nella definizione di “banca dati”, come ivi prevista.

Deve ora osservarsi che, invece, la parola “mining” utilizzata dal legislatore europeo nella CDSMD nella formula TDM, seppure tradotta anch’essa nella versione italiana della CDSMD con la parola “estrazione”, riguarda un’attività totalmente diversa, ossia una tecnica di “analisi automatizzata” di “dati in formato digitale”, come si ricava chiaramente dalla relativa definizione.

La parola “mining” (in inglese) è una parola che il legislatore europeo ha tratto dalla prassi, ed è riconosciuto che si tratti di una parola impiegata nel campo della scienza informatica più per la sua forza suggestiva che per la sua capacità di evocare un campo di significato pregnante e coerente all’attività che si intende con essa designare e che consiste – si ribadisce – in una “analisi automatizzata” di testi e di dati in formato digitale «avente lo scopo di generare informazioni inclusi, a titolo non esaustivo, modelli, tendenze e correlazioni». Pertanto, ai fini dell’analisi e della rilevanza giuridica, non c’è dubbio alcuno che non si possa confondere l’estrazione di cui alla direttiva banche dati 96/9/CE con l’estrazione (*mining*) dell’acronimo TDM di cui alla CDSMD, perché in quest’ultimo contesto la parola estrazione individua una attività ed una tecnica di analisi automatizzata di dati in formato digitale aventi un determinato scopo.

Sulla base di ciò, e di tutte le altre considerazioni svolte in questo paragrafo e nei paragrafi precedenti, possiamo perciò senza alcun dubbio dire che l’estrazione è unicamente ed esclusivamente rilevante, come attività riservata, nel sistema degli artt. 3 e 4 CDSMD, relativamente al diritto di estrazione delle banche dati ai sensi della omonima direttiva 96/9/CE. Ed aggiungere con fermezza che la CDSMD né presuppone né tanto meno prevede *ex novo* un diritto esclusivo di estrazione di dati relativi ad opere e altri materiali protetti. Un simile diritto non esiste, come detto, nel diritto europeo, nel senso che non si trova riconosciuto o creato in alcun’altra norma del diritto euro-unitario.

7. Relativamente alla sola disciplina italiana di recepimento della CDSMD (art. 70-ter, co. 1, l.a.): l’inesistenza di un generale divieto di comunicazione al pubblico dei risultati delle ricerche compiute in esito all’applicazione di tecniche di TDM

Rivolgendoci poi più in particolare alla disciplina italiana di recepimento della CDSMD, segnaliamo che

un’altra regola inesistente che rischia di essere creata attraverso un’erronea interpretazione dell’art. 70-ter co. 1 l.a., attesa la singolare formulazione dell’ultimo periodo di tale disposizione⁴⁵ – priva di riscontro nella CDSMD – è che esisterebbe un generale divieto di comunicazione al pubblico dei risultati delle ricerche compiute in esito all’applicazione di tecniche di TDM.

Non è possibile dilungarsi in questa sede sulle significative e gravi conseguenze sistematiche ed applicative che dovrebbero trarsi laddove dovesse ritenersi esistente un simile (invece inesistente) divieto.

Qui è sufficiente osservare: (i) che la CDSMD non contiene alcuna disposizione del tenore di quella dell’art. 70-ter l.a.; (ii) che l’ordinamento italiano non contiene alcuna disposizione che possa interpretarsi nel senso di supportare l’esistenza di un simile divieto.

8. Il sillogismo alla base della c.d. eccezione di TDM

Proseguendo nell’analisi, e chiariti i confini del “diritto di estrazione” (*extraction*, non: *mining*) al quale si riferisce il sistema delle eccezioni e limitazioni di cui agli artt. 3 e 4 CDSMD, si evidenzia la necessità di delineare anche i confini del “diritto di riproduzione”, al quale si riferiscono i medesimi articoli. In proposito, si osserva preliminarmente come la c.d. eccezione TDM disegnata dalla CDSMD sia originata da un sillogismo, logicamente strutturato (come ogni sillogismo) in tre elementi: premessa maggiore, premessa minore, conclusione.

Precisamente, la premessa maggiore è la seguente: la riproduzione e l’estrazione da opere e altri materiali protetti menzionati negli artt. 3 e 4 CDSMD sono attività riservate *ex lege* ai titolari di omonimi diritti esclusivi ai quali è accordata la relativa protezione con la conseguenza che i terzi, per svolgere tali attività, o devono beneficiare di una eccezione di legge oppure devono soggiacere ad un potere di autorizzazione o divieto, *breviter* di licenza, dei titolari dei diritti, alle condizioni previste dalla legge.

La premessa minore è la seguente: le attività tecniche poste in essere ai fini delle operazioni di TDM comprendono attività di riproduzione o di estrazione riservate *ex lege*.

La conclusione è la seguente: per consentire le attività tecniche poste in essere ai fini delle operazioni di TDM c’è bisogno o di una eccezione di legge o di una licenza nelle forme di legge⁴⁶.

Va in proposito segnalata la diversa natura delle due premesse: la premessa maggiore avendo la natura di una statuizione normativa, quella minore la natura di statuizione fattuale.



Da ciò – osserviamo – si ricava anche che la premessa minore non possa avere validità generale e che, per forza di logica, debba predicarsi – ovvero non possa escludersi – la possibilità della ricorrenza di casi concreti in cui le attività tecniche poste in essere ai fini delle operazioni di TDM non siano qualificabili come attività di riproduzione o di estrazione in senso tecnico, riservate *ex lege* (per il possibile difetto in concreto degli elementi di fatto che consentono una simile qualificazione in punto di diritto)⁴⁷, con conseguente impossibilità, in tali casi, di trarre la conclusione del sillogismo. In questo senso, si deve logicamente negare che il TDM comporti di per sé, ossia *necessariamente*, attività di riproduzione o estrazione riservate *ex lege* dalla legislazione euro-unitaria.

9. Non è vero che il TDM comporti *necessariamente* attività di riproduzione o estrazione riservate *ex lege* nella legislazione europea

Da quanto sopra, deduciamo che un'altra regola inesistente che rischia di essere creata attraverso l'eccezione se non si interpreta correttamente la nuova disciplina della CDSMD è quella per cui il TDM comporti sempre un'attività di estrazione o di riproduzione riservata *ex lege* e debba di conseguenza essere sempre (ossia *necessariamente*) assoggettato al vaglio di legittimità previsto dal regime di diritto euro-unitario delle eccezioni e limitazioni di cui agli artt. 3 e 4 della CDSMD. Al contrario, osserviamo come sia corretto dichiarare che un simile assoggettamento debba valutarsi *di volta in volta*, sulla base di un'analisi delle fattispecie concrete diretta a verificare in ciascun caso se le attività poste in essere ai fini dell'analisi automatizzata dei dati digitali chiamata "TDM" comprendano o meno attività rispondenti ai precisi significati tecnico-giuridici che attraverso le parole "estrazione" e "riproduzione" identificano quelle specifiche attività così nominate e riservate *ex lege* dal diritto euro-unitario ai titolari dei diritti esclusivi ai quali si riferiscono gli artt. 3 e 4 della CDSMD.

10. Il concetto di riproduzione e i dati leggibili solo dagli elaboratori elettronici

La successiva domanda che si propone in conseguenza della suesposta analisi è se il TDM comporti, *se non necessariamente*, quantomeno *normalmente*, ossia ordinariamente (alla stregua di un criterio di regolarità statistica o di ricorrenza prevalente in concreto) attività di riproduzione o estrazione riservate *ex lege*. Ed

infatti, osserviamo che se questo fosse il caso potrebbe spiegarsi quantomeno la *ratio* della nuova disciplina in commento (impregiudicata, s'intende, ogni altra considerazione critica relativa alla tecnica normativa utilizzata nella CDSMD). Ciò conferma la necessità che – dopo la premessa di carattere logico sulla struttura del sillogismo presupposto nel regime della eccezione di TDM introdotto dalla CDSMD – si prenda posizione anche sul significato tecnico-giuridico della parola "riproduzione".

Rileviamo in proposito che, come per la parola "estrazione", anche per quella "riproduzione" deve ricercarsi con rigore il significato giuridicamente rilevante nel contesto degli artt. 3 e 4 CDSMD, e non anche qualunque significato rimandato dal linguaggio comune. Nel citato studio del gennaio 2022 della Commissione europea su copyright e nuove tecnologie, effettivamente, ci si concentra sull'attività di riproduzione e si espone il dubbio che le attività strumentali al TDM siano tali da violare sempre il diritto di riproduzione, come contemplato dalla legislazione euro-unitaria rilevante ai sensi degli artt. 3 e 4 CDSMD. A tal proposito, nel citato studio si afferma essere rilevante stabilire se l'opera protetta sia riconoscibile (*recognisable*) e dunque sfruttata (*exploited*) nel corso del processo di TDM o successivamente, nel senso che solo in questo caso possa predicarsi un'interferenza potenzialmente lesiva del diritto di riproduzione in assenza di elementi di fatto integranti un'eccezione o una licenza ai sensi degli artt. 3 e 4 CDSMD.

Sviluppando questa linea di pensiero, si reputa ragionevole, adeguato alla realtà tecnologica del tempo presente e giuridicamente corretto (dal punto di vista della tecnica e della logica giuridica), sostenere che l'analisi automatizzata di dati *non* integra di per sé una attività di riproduzione nel senso tecnico-giuridico inteso dalle norme che contemplano tale diritto esclusivo nella legislazione euro-unitaria, laddove essa abbia ad oggetto dati in formato digitale esclusivamente *machine readable* (ossia dati leggibili solo da elaboratori elettronici)⁴⁸.

Tale conclusione si basa sulla *inidoneità di un simile processo a creare di per sé una riproduzione di forme espressive*⁴⁹ *intelligibili dall'uomo* delle opere e degli altri materiali protetti e un conseguente loro sfruttamento *come tali*.

In questa prospettiva, va sollecitata e promossa un'interpretazione adeguata ai tempi delle norme della direttiva InfoSoc (Direttiva 2001/29/CE sull'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione) in particolare per quanto riguarda il *wording* dell'art. 2 «Gli Stati membri riconoscono ai soggetti sotto elencati il diritto esclusivo di autorizzare o vietare la



riproduzione diretta o indiretta, temporanea o permanente, in qualunque modo o forma, in tutto o in parte [...]»⁵⁰, anche alla luce dei precedenti della Corte di Giustizia della UE citati nel predetto studio del 2022 della Commissione europea (v. *retro*, paragrafo 3) a supporto dell'interpretazione che anche qui si sostiene, tra cui il precedente del caso *Pelham* del 2019⁵¹.

11. Non è vero che il TDM comporti nemmeno ordinariamente attività di riproduzione o estrazione riservate *ex lege* nella legislazione europea

Sulla base delle osservazioni di cui sopra, può desumersi che il TDM non comporta nemmeno normalmente attività di riproduzione o estrazione riservate *ex lege* nella legislazione euro-unitaria.

Ed infatti, considerato che i processi di TDM lavorano per definizione su dati digitali o su copie di dati digitali, e che in entrambi i casi si tratta ordinariamente di dati esclusivamente *machine readable*, si fa discendere come ulteriore conseguenza l'affermazione che il TDM non comporta *nemmeno ordinariamente* attività di riproduzione o estrazione riservate *ex lege* ai sensi della legislazione euro-unitaria rilevante.

12. Le tendenze del nuovo diritto europeo sui dati e la diversa interpretazione della CDSMD in funzione di protezione dei diritti di proprietà intellettuale

L'erigendo diritto euro-unitario dei dati⁵² presenta la netta tendenza a vietare espressamente certi usi (dei dati digitali) identificati per la loro finalità tipica o per la loro oggettiva idoneità a ledere diritti dei terzi⁵³.

Questa osservazione – che riguarda il piano dell'utilizzazione dei dati, le sue finalità ed oggettive potenzialità – serve a individuare soluzioni interpretative che corrispondono a quell'esigenza di tutela sottesa alla disciplina del TDM, permettendo di respingere la non corretta interpretazione che postula l'esistenza di un inesistente diritto di TDM, ed al contempo di assicurare protezione ai diritti che la CDSMD intende proteggere.

L'interprete può ed anzi deve senz'altro ricavare dal sistema del nuovo diritto euro-unitario dei dati anche un divieto di utilizzare processi di TDM per fini, o con modalità oggettivamente idonee a comportare ipotesi di plagio e/o di contraffazione di opere o altri materiali protetti dal diritto d'autore o altrimenti la

violazione di diritti esclusivi e dei diritti connessi. Un simile divieto (che può senz'altro ricavarsi dal sistema) deve ritenersi sufficiente a tutelare gli interessi protetti dal diritto euro-unitario sul copyright e i diritti connessi nel mercato unico digitale. Al contrario, un contrasto *a priori* dei processi di TDM costituirebbe un sicuro fattore di ostacolo e di depressione della ricerca e dell'economia nel campo dell'intelligenza artificiale (ivi compresa l'economia della c.d. *creative industry*), e comporterebbe ermeneuticamente un'ingiustificata espansione dei diritti di proprietà intellettuale con una, speculare, ingiustificata e sproporzionata compressione di diritti fondamentali riconosciuti dal diritto euro-unitario (artt. 11, 13, 15 e 16 Carta dei diritti fondamentali dell'Unione europea).

13. Il TDM come risorsa (invece che pericolo) per l'industria del copyright

Una volta messo a fuoco il predetto divieto (v. paragrafo precedente), si può comprendere come il TDM, lungi dal costituire un pericolo per l'industria del copyright, ne costituisce effettivamente una risorsa, essenziale ed imprescindibile nella realtà contemporanea.

Sarebbe irrealistico, oltre che giuridicamente errato, bloccare l'industria dell'intelligenza artificiale nel senso inteso dalla (inesatta) postulazione di un (inesistente) diritto di TDM.

Sarebbe anche controproducente per la stessa "industria del copyright".

L'industria del copyright ha reali prospettive di guadagno da programmi software e autori che sviluppano e impiegano tecniche di TDM. Non a caso il Governo del Regno Unito ha deciso di procedere nel senso sopra illustrato (v. *retro*, paragrafo 2).

Il problema è piuttosto quello di stabilire il confine tra un uso lecito ed un uso illecito del TDM. Ma questo non può davvero scoraggiare gli interpreti e gli operatori del mondo del copyright, i quali sono ben avvezzi a riconoscere la mobilità e la labilità dei confini tra pratiche di plagio ed illecite e pratiche lecite. Si tratta di cimentarsi con una nuova materia (per stabilire di volta in volta il confine tra un uso lecito ed un uso illecito del TDM), ma la manualità e il mestiere (le categorie e i principi) non cambiano.

14. Conclusioni

La nuova disciplina degli artt. 3 e 4 CDSMD rischia, se non correttamente e rigorosamente inquadrata, di creare ermeneuticamente e surrettiziamente un nuovo



sistema di IPRs sui dati, che la stessa struttura logica delle eccezioni e delle limitazioni relative a diritti che hanno ad oggetto beni immateriali diversi dai dati impone senz'altro di escludere.

Al contrario, va affermato con chiarezza che gli artt. 3 e 4 CDSMD non presuppongono né costituiscono alcun diritto esclusivo di TDM. Per i motivi che abbiamo provato a riassumere nel presente intervento, riteniamo che non soltanto per il diritto internazionale d'autore⁵⁴, ma anche per il diritto euro-unitario, sia sbagliato ritenere esistente *di per sé* la questione della violazione del diritto d'autore e dei diritti connessi da parte delle attività di TDM.

Anche per il diritto euro-unitario, l'interferenza del TDM con i regimi di esclusiva si pone in relazione alle attività di riproduzione e di estrazione (quest'ultima limitatamente al diritto *sui generis* sulle banche dati) *soltanto se e nella misura in cui* nel contesto delle operazioni di analisi automatizzata di dati digitali, in che consiste il TDM, vengano poste in essere attività di riproduzione ed estrazione precisamente ed esclusivamente intese nel modo statuito dalle normative che le riservano ai titolari dei diritti di esclusiva.

Si tratta dunque di una interferenza solo eventuale e non necessaria.

Per i motivi che abbiamo provato ad illustrare nel presente articolo, deve inoltre ritenersi che il trattamento automatizzato di dati digitali esclusivamente *machine readable* non comporti *di per sé* alcuna interferenza con i regimi di esclusiva. Per cui, operando la tecnica del TDM normalmente su dati digitali esclusivamente *machine readable*, l'interferenza non è nemmeno ordinaria.

Al legislatore europeo della CDSMD va mossa una forte critica di tecnica legislativa, perché l'aver creato quel sistema di eccezioni e limitazioni (artt. 3 e 4 CDSMD) comunica superficialmente al lettore della direttiva un messaggio sbagliato, ossia comunica l'idea che un problema di interferenza tra le attività di TDM e i diritti sulle opere protette dal diritto di autore esista sempre. Questo, per il motivo sopra riassunto, è sbagliato, con la conseguenza pratica che tutti i moduli e gli standard di contratto di licenza già ampiamente diffusi sul mercato che hanno ad oggetto "diritti TDM"⁵⁵ non hanno una base giuridica *come tali*, perché non esistono diritti esclusivi di TDM che possano licenziarsi.

Il problema della protezione dei diritti di proprietà intellettuale considerati dagli artt. 3 e 4 CDSMD si pone su un diverso piano, che è quello proprio del nuovo diritto europeo dei dati, ossia il piano della potenziale lesione dei diritti dei terzi (in questo caso diritti di proprietà intellettuale) da parte di certi usi di dati digitali. Ma questo – che è un problema di

finalità ed idoneità di certi usi e relativi output a ledere certe espressioni tutelate dal regime del copyright generalmente inteso – è un problema antico, che non può risolversi con la scorciatoia di postulare generalmente un diritto esclusivo di TDM: il quale, a ben vedere, non esiste nel diritto euro-unitario e tanto meno è stato creato dalla CDSMD.

L'industria del copyright ha tutto da guadagnare dal TDM attraverso lo sviluppo di nuovi software e il loro impiego da parte di autori tecnologicamente innovativi. I professionisti della protezione del copyright devono lavorare nella suddetta dimensione della tutela (quella che li chiama a stabilire di volta in volta il confine tra usi leciti ed illeciti delle tecnologie di TDM) che è sicuramente più impegnativa della dimensione da alcuni erroneamente intravista nella scorciatoia di un inesistente diritto di TDM. Ma trarranno beneficio anch'essi dai maggiori guadagni assicurati da questa nuova tecnologia.

Note

¹Direttiva (UE) 2019/790 del Parlamento europeo e del Consiglio del 17 aprile 2019 sul diritto d'autore e sui diritti connessi nel mercato unico digitale e che modifica le direttive 96/9/CE e 2001/29/CE.

²Art. 3 CDSMD: «Estrazione di testo e di dati per scopi di ricerca scientifica – 1. Gli Stati membri dispongono un'eccezione ai diritti di cui all'articolo 5, lettera a), e all'articolo 7, paragrafo 1, della direttiva 96/9/CE, all'articolo 2 della direttiva 2001/29/CE, e all'articolo 15, paragrafo 1, della presente direttiva per le riproduzioni e le estrazioni effettuate da organismi di ricerca e istituti di tutela del patrimonio culturale ai fini dell'estrazione, per scopi di ricerca scientifica, di testo e di dati da opere o altri materiali cui essi hanno legalmente accesso. 2. Le copie di opere o altri materiali realizzate in conformità del paragrafo 1 sono memorizzate con un adeguato livello di sicurezza e possono essere conservate per scopi di ricerca scientifica, inclusa la verifica dei risultati della ricerca. 3. I titolari dei diritti sono autorizzati ad applicare misure atte a garantire la sicurezza e l'integrità delle reti e delle banche dati in cui sono ospitate le opere o altri materiali. Tali misure non vanno al di là di quanto necessario per il raggiungimento di detto obiettivo. 4. Gli Stati membri incoraggiano i titolari dei diritti, gli organismi di ricerca e gli istituti di tutela del patrimonio culturale a definire concordemente le migliori prassi per l'applicazione dell'obbligo e delle misure di cui rispettivamente ai paragrafi 2 e 3».

³Art. 4 CDSMD: «Eccezioni o limitazioni ai fini dell'estrazione di testo e di dati - 1. Gli Stati membri dispongono un'eccezione o una limitazione ai diritti di cui all'articolo 5, lettera a), e all'articolo 7, paragrafo 1, della direttiva 96/9/CE, all'articolo 2 della direttiva 2001/29/CE, all'articolo 4, paragrafo 1, lettere a) e b), della direttiva 2009/24/CE e all'articolo 15, paragrafo 1, della presente direttiva per le riproduzioni e le estrazioni effettuate da opere o altri materiali cui si abbia legalmente accesso ai fini dell'estrazione di testo e di dati. 2. Le riproduzioni e le estrazioni effettuate a norma del paragrafo 1 possono essere conservate per il tempo necessario ai fini dell'estrazione di testo e di dati. 3. L'eccezione o la limitazione di cui al paragrafo 1 si applica a condizione che l'utilizzo delle opere e di altri materiali di cui a tale paragrafo non sia stato



espressamente riservato dai titolari dei diritti in modo appropriato, ad esempio attraverso strumenti che consentano lettura automatizzata in caso di contenuti resi pubblicamente disponibili online. 4. Il presente articolo non pregiudica l'applicazione dell'articolo 3 della presente direttiva».

⁴Nell'impianto dell'art. 4 CDSMD, l'autorizzazione da parte dei titolari dei diritti si reputa fornita *ex lege* in assenza di un loro esplicito diniego o "riserva", c.d. *opt-out*, e sempre che si tratti di materiali ai quali i terzi hanno "legalmente accesso". Nella recente consultazione pubblica promossa dal Governo del Regno Unito, su cui vedi *infra* nel testo, lo schema di tale soluzione era tra le quattro opzioni considerate (opzione 3). Il Governo del Regno Unito ha annunciato che procederà invece nella direzione dell'eliminazione di qualunque restrizione all'eccezione di TDM (opzione 4).

⁵*Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation.*

⁶Cfr. T. MARGONI, *The new EU Copyright in the Digital Single Market Directive (CDSMD) and its impact on post-Brexit UK researchers*, June 2020. Una interessante analisi comparativa, che giunge a conclusioni in linea con quelle argomentate nel presente contributo, si trova svolta in M. SENFTLEBEN, *Compliance of National TDM Rules with International Copyright Law – An Overrated Nonissue?*, in "International Review of Intellectual Property and Competition Law" (IIC), vol. 53, 2022, n. 10, p. 1477 ss.

⁷Cfr. sul punto A. OTTOLIA, *L'utilizzo computazionale dell'opera dell'ingegno in internet*, in "AIDA", 2014, p. 405 s.

⁸Punto 31 del Documento del 28 giugno 2022: «Text and data mining (TDM) means using computational techniques to analyse large amounts of information to identify patterns, trends and other useful information».

⁹Per uno studio comparativo e un approfondimento sul diritto internazionale d'autore, v. M. SENFTLEBEN, *Compliance of National TDM Rules with International Copyright Law*, cit. In questo lavoro, l'a. osserva che, da una analisi di diritto comparato, risultano approcci differenti al governo delle attività di TDM. L'a. commenta che, mentre uno spazio notevole per il TDM può conseguire dall'applicazione della dottrina statunitense del *fair use* così come dalle ampie esenzioni del tipo di quelle introdotte in Giappone, gli Stati dell'Unione europea si affidano a una regolamentazione maggiormente restrittiva basata su specifiche eccezioni, che produce una notevole complessità e genera incertezza giuridica anche per la circostanza di dipendere da criteri (come quello delle finalità di ricerca scientifica e quello dell'assenza di una riserva di diritti da parte dei titolari dei diritti) la cui disponibilità deve essere di volta in volta verificata su base individuale. L'a. si interroga anche sulla possibilità che i legislatori possano ritenersi vincolati a seguire un approccio restrittivo alla luce del diritto internazionale, in particolare con riferimento al "three-step test" di cui agli artt. 9(2) della Convenzione di Berna, 3 del *TRIPS Agreement* e 10 del *WIPO Copyright Treaty*. Tuttavia, l'a. esclude che il diritto internazionale richieda un approccio restrittivo, ed anzi mette in dubbio che il diritto internazionale copra in assoluto le attività di TDM. L'a. nota innanzitutto, a tal riguardo, che il TDM non è ascrivibile alla tradizionale categoria di uso che può essere stata considerata nelle conferenze diplomatiche che hanno preparato i testi della Convenzione di Berna, del *TRIPS Agreement* o del *WIPO Copyright Treaty*. Il TDM – osserva correttamente l'a. – consiste in un uso analitico ed automatizzato che non tocca in alcun modo l'*expressive core* delle opere letterarie ed artistiche. Da ciò l'a. argomenta – condivisibilmente – la conclusione che il TDM si colloca completamente al di fuori dell'ambito di armonizzazione del diritto internazionale d'autore. Dalla prospettiva così tracciata, l'a. osserva correttamente che la discussione sulla conformità delle attività di TDM alle norme di diritto internazionale d'autore è

priva di senso, e che – di conseguenza – il diritto internazionale d'autore non limita la libertà dei legislatori nazionali di disegnare adeguate soluzioni di diritto nazionale per riconciliare la protezione del diritto d'autore con il diritto di ricerca che è sotteso alle attività di TDM. Per i motivi che illustreremo nel prosieguo del presente lavoro, non soltanto condividiamo l'analisi svolta da M. Senftleben (nel sopra citato contributo), ma riteniamo anche corretto sostenere che la stessa conclusione da egli formulata sulle fonti di diritto internazionale d'autore, i.e. la non interferenza delle attività di TDM di per sé considerate con i regimi di esclusiva del diritto d'autore e dei diritti connessi, debba trarsi anche per il diritto euro-unitario.

¹⁰EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR COMMUNICATION, *Study on copyright and new technologies: copyright data management and artificial intelligence*, Publications Office of the European Union, 2022.

¹¹*Ivi*, p. 110 ss., spec. p. 192 ss. Per una serrata critica al legislatore europeo, in termini di perdita di competitività dell'industria europea dell'IA indotta dalla disciplina del TDM di cui agli artt. 3 e 4 CDSMD cfr. C. GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, PIJIP/TLS Research Paper Series, 2021, n. 66, in M. Senftleben, J. Poort, M. van Eechoud et al. (eds.), "Intellectual Property and Sports, Essays in Honour of P. Bernt Hugenholtz", Kluwer Law International, 2021, pp. 383-394.

¹²Un'espressa, condivisibile e motivata critica si trova svolta in T. MARGONI, M. KRETSCHMER, *A deeper look into the EU Text and Data Mining exceptions: Harmonisation, data property and the future of technology*, in CREATE Working Paper 2021, n. 7, v. in particolare nella sintesi che si trova a pp. 32-33. Sul ruolo della WIPO con specifico riferimento ai problemi di intersezione tra intelligenza artificiale e TDM, v. anche S. FLYNN, C. GEIGER, J.P. QUINTAIS et al., *Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action*, in "European Intellectual Property Review", vol. 42, 2020, n. 7, pp. 393-398, WCL Research Paper n. 2020-12. La proposta di CDSMD, in particolare quanto all'eccezione di TDM, veniva opportunamente analizzata nella prospettiva della congruenza con l'economia dei dati da M.L. MONTAGNANI, G. AIME, *Il text and data mining e il diritto d'autore*, in "AIDA", 2017, p. 376 ss.

¹³EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR COMMUNICATION, *Study on copyright*, cit., p. 215.

¹⁴*Ibidem*, «Lastly, the argument according to which, the introduction of the TDM exceptions in the DSM dir. implies that the technical acts needed for mining some content are falling under the reproduction right can be challenged. Some legal exceptions do only confirm the boundaries of the rights or the state of the law, and are expressly recognized in the law to increase the security for the market participants and to codify and clarify the law as it applied. Even if the exception can be set aside by an opt-out, this does not entail that all technical copies made in the TDM process are to be considered as reproductions: the opt-out is meant to reserve the right, where the right is applicable, i.e. where a 'reproduction' can be found (in particular determining whether the work is recognisable and thus exploited in the course of the TDM process or afterwards) [...] Along this reasoning, a developer or user of an AI solution could argue that the technical copies made of protected cultural content in the course of a TDM process (while training an AI solution) are not sufficient for finding an infringement of the reproduction right».

¹⁵Tra gli altri A. STROWEL, *Reconstructing the reproduction and the communication the public rights: how to align copyright with its fundamentals*, in P. Bernt Hugenholtz (ed.), "Copyright Reconstructed Rethinking Copyright's Economic Rights in a



Time of Highly Dynamic Technological and Economic Change”, Wolters Kluwer, 2018, p. 203; S. DEPREEUW, *The Variable Scope of the Exclusive Economic Rights in Copyright*, Kluwer, 2014, p. 189 ss. e bibliografia *ivi* citata. V. anche R. DUCATO, A. STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, in CRIDES Working Paper Series no. 1/2021, p. 24 ss.

¹⁶CGUE 29 luglio 2019, *Pelham GmbH e altri c. Ralf Hütter e Florian Schneider-Esleben*, C-476/1, caso nel quale la Corte di Giustizia ha negato che ricorresse una riproduzione nel caso in cui un musicista, esercitando la propria libertà delle arti, aveva prelevato un campione sonoro da un fonogramma altrui al fine di integrarlo, in forma modificata e non riconoscibile all’ascolto umano, in un nuovo fonogramma.

¹⁷EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR COMMUNICATION, *Study on copyright*, cit., p. 213: «The vast scope of protection of the reproduction, covering ephemeral, invisible copies of a purely technical nature [...] can be explained by historic reasons and has been criticised in scholarly literature. Especially for the ephemeral and partial reproductions, the literal and technical reading equates every ‘copy’ (in the technical sense) to a ‘reproduction’ (in the legal sense), without considering the importance of the ‘copy’ for the exploitation of the ‘work’. Arguably, the notion of ‘reproduction’ could be revisited to align its content with the purpose of the economic rights, i.e. protect the exploitation of the work or other protected subject matter. To the extent that the technical copies made in the course of a TDM process serve to extract information, meta-data or other insights from the processed training data set, the protected creations in the data set are used for the sake of the information they contain, not for the sake of commercialising, exploiting or putting forward the protected works in their original expression. It could be argued then that the contours of the reproduction right should be revisited to focus its protection on the protection of the works (and other subject matter) and their exploitation of the protected aspects (i.e. the original expression, the recognisable performance, the recognisable recordings). Although the decisions of the CJEU on the reproduction right have oftentimes put forward a technical and literal reading of the reproduction right, the Court has demonstrated an openness to a more flexible and more balanced approach of the reproduction right in the *Pelham* decision. In *Pelham*, the Court contrasted a literal interpretation (which would be warranted by the purposes of providing a high level of protection and protecting the producer’s investment) to a contextual interpretation (which would seek a fair balance with other fundamental rights) of the notion ‘reproduction in part’ of a phonogram».

¹⁸Per una approfondita analisi delle questioni inerenti l’eccezione di TDM e la ricerca scientifica nel disegno della CDSMD v. R. CASO, *Il conflitto tra diritto d’autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, Trento LawTech Research Paper n. 38, 2020. Un inquadramento generale della direttiva è offerto in L. CASALINI, *Accesso e digitalizzazione nella direttiva copyright*, in “Persona e mercato”, 2021, n. 2, p. 412 ss., ove anche interessanti considerazioni sulle possibilità di sviluppo aperte da un corretto processo di digitalizzazione. Per una disamina delle questioni del TDM già durante la proposta di CDSMD, cfr.: M.L. MONTAGNANI, G. AIME, *Il text and data mining*, cit.; gli studi commissionati dalla Commissione JURI del Parlamento europeo, in particolare E. ROSATI, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects*, 2018; C. GEIGER, G. FROSIO, O. BULAYENKO, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects*, 2018; ID., *Text and Data Mining in the Proposed Copyright*

Reform: Making the EU Ready for an Age of Big Data?, in “International Review of Intellectual Property and Competition Law”, vol. 49, 2019, n. 7, p. 814 ss.

¹⁹Cfr. *ex multis* T. MARGONI, M. KRETSCHMER, *A deeper look into the EU Text*, cit., p. 7: «Not only TDM *stricto sensu* has been limited to research uses by research and cultural organisations, but virtually any automated technique that analyses information in digital form is captured under the narrow boundaries of the current formulation of Arts. 3 and 4. This certainly includes most modern, data-driven forms of AI, such as traditional machine learning and more advanced forms of deep learning and neural network structures. The policy reasons justifying the allocation of the power to authorise these cutting-edge technologies to upstream players in the database and content markets are far from clear».

²⁰Cfr. T. MARGONI, M. KRETSCHMER, *A deeper look into the EU Text*, p. 33: «An overall assessment of the situation portrayed in this paper cannot be optimistic. Whereas a good amount of attention in scholarship has been (rightly) dedicated to critically evaluate recent proposals to create a data producer right, this paper shows that the EU legislator, probably even beyond its own intentions, has taken a very drastic position on a complementary and highly relevant matter, the ownership of pre-existing mere facts and data. As demonstrated, this position is not functional to a proportionate, fair, and accountable regulatory framework for copyright, for technology and for the EU as an economic, social and political institution».

²¹Al momento constano licenze standard (*TDM license*) preparate dagli editori, e non anche da parte di organismi di gestione collettiva. V. ad esempio: *Elsevier Text and Data Mining (TDM) License*; *Text and Data Mining on SAGE Journals*; *TDM Licence/Policy - BMJ Text and Data Mining (TDM) Policy and Licence*; *ASM Journals Non-Commercial TDM License*. Per un modello di *web protocol* espressivo della riserva dei diritti di cui all’art. 4 CDSMD v. W3C COMMUNITY GROUP, *TDM Reservation Protocol (TDMRep) Final Community Group Report*.

²²Ossia il diritto di riproduzione previsto dall’articolo 5, lettera a) della direttiva banche dati 96/9/CE, il diritto di estrazione previsto dall’articolo 7, paragrafo 1, della direttiva banche dati 96/9/CE, il diritto di riproduzione previsto dall’articolo 2 della direttiva InfoSoc 2001/29/CE, il diritto di riproduzione previsto dall’articolo 4, paragrafo 1, lettere a) e b), della direttiva software 2009/24/CE, e il diritto di riproduzione previsto dall’articolo 15, paragrafo 1, della medesima CDSMD (utilizzo online delle pubblicazioni giornalistiche).

²³Riconosce come «in linea di principio, la disciplina del diritto d’autore non conferisce alcun controllo sui dati» P. GUARDA, *Il regime giuridico dei dati della ricerca scientifica*, 2021, p. 100. Cfr. A. OTTOLIA, *Big Data e innovazione computazionale*, in “Quaderni di AIDA”, 2017, n. 28, pp. 22 ss., 163 ss., 271 ss.; ID., *L’utilizzazione computazionale*, cit., *passim*, spec. p. 405 ss.; M.L. MONTAGNANI, G. AIME, *Il text and data mining*, cit.; G. ROSSI, *Opere dell’ingegno come dati: il text e data mining nella direttiva 2019/790*, in “AIDA”, 2019, pp. 235-256, osserva che le attività di *text and data mining* utilizzano le opere protette dal diritto d’autore come meri dati, ossia senza interferire con il loro effettivo valore di espressione creativa di idee. L’a. nota che questa era stata la questione centrale nel famoso contenzioso (negli USA) *Authors’ Guild v. Google e Authors’ Guild v. Hathi Trust*, in esito al quale le corti statunitensi hanno dichiarato che la digitalizzazione massiva delle opere letterarie per fini di *data mining* costituisce un *transformative use*, come tale coperto dalla dottrina del *fair use*, a ciò non ostando gli eventuali scopi commerciali di Google. L’a. prosegue affermando che, al contrario, la CDSMD ha distinto le attività di TDM poste in essere da organismi di ricerca e istituti di tutela del patrimonio culturale,



da un lato, da attività simili poste in essere per finalità diverse, incluse finalità commerciali, dall'altro lato. L'a. prosegue commentando che questo diverso trattamento (oltre a generare significativi problemi interpretativi) sembra non essere adeguato rispetto alla realtà e alla natura della ricerca algoritmica. Interessanti osservazioni sono infine contenute in questo saggio in tema di comparazione tra le tecniche di protezione (giuridica) dei dati personali e delle opere protette dal diritto d'autore con riferimento ai "big data"; C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and Data Mining: articles 3 and 4 of the directive 2019/790/EU*, CEIPI (Centre for International Intellectual Property Studies Research) paper n. 8/2019; R. CASO, *Il conflitto tra diritto d'autore e ricerca scientifica*, cit., spec. p. 15 ss., dove l'a. nota le ambiguità del testo della CDSM e osserva che la direttiva offre una definizione di TDM ma non anche una definizione di "dato", "testo" e "informazione"; L. MANSANI, *Le eccezioni per estrazione di testo e dati, didattica e conservazione del patrimonio culturale*, in "AIDA", 2019, p. 3 ss., osserva che la disciplina della CDSMD sul TDM è di grande impatto sia da un punto di vista economico che per i profili teorici implicati. L'a. afferma che le soluzioni normative di cui alla medesima direttiva sembrano il risultato di un compromesso tra i titolari dei diritti e gli utilizzatori delle opere protette, in alcuni casi con uno sbilancio a favore dei primi, avuto riguardo, in particolare, alle soluzioni in vigore in altri ordinamenti; A. MUSSO, *Eccezioni e limitazioni al diritto d'autore nella Direttiva UE n. 2019/790*, in "Diritto dell'informazione e dell'informatica", 2020, n. 3, p. 411 ss. (Relazione al convegno "Direttiva Copyright, le sfide del recepimento in Italia", Roma, Biblioteca Casanatense, 11 settembre 2019); J.P. TRIAILLE et al., *Study on the Legal Framework of Text and Data Mining (TDM)*, 2014.

²⁴Uno studio del 2022 richiesto dalla speciale commissione AIDA (*Artificial Intelligence in a Digital Age*) del Parlamento europeo registrava ben 48 atti normativi di diritto euro-unitario (tra quelli emanati e da emanare) in campo digitale: C. CODAGNONE, G. LIVA, T. RODRIGUEZ DE LAS HERAS BALLELL, *Identification and assessment of existing and draft EU legislation in the digital field*.

²⁵Direttiva (UE) 2019/1024 del Parlamento europeo e del Consiglio del 20 giugno 2019 relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico.

²⁶Regolamento (UE) 2022/868 del Parlamento europeo e del Consiglio del 30 maggio 2022 relativo alla governance europea dei dati e che modifica il regolamento (UE) 2018/1724.

²⁷Proposta di regolamento del Parlamento europeo e del Consiglio del 23 febbraio 2022, COM(2022) 68, riguardante norme armonizzate sull'accesso equo ai dati e sul loro utilizzo (normativa sui dati).

²⁸Proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, COM(2021) 206, che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione.

²⁹Art. 1, par. 2 lett. c) Direttiva *Open Data* (UE) 2019/1024: «La presente direttiva non si applica: [...] c) ai documenti su cui terzi detengono diritti di proprietà intellettuale».

³⁰Art. 2, n. 1) Regolamento *Data Governance Act* (UE) 2022/868: «"dati": qualsiasi rappresentazione digitale di atti, fatti o informazioni e qualsiasi raccolta di tali atti, fatti o informazioni, anche sotto forma di registrazione sonora, visiva o audiovisiva».

³¹Nei considerando 10, 12 e 20 del Regolamento *Data Governance Act* (UE) 2022/868 si trova ben formulata la distinzione tra i dati e i beni immateriali che formano – questi ultimi soltanto – oggetto di diritti di proprietà intellettuale («10. Le categorie di dati detenuti da enti pubblici, che dovrebbero essere soggetti al riutilizzo a norma del presente regolamento, non rientrano nell'ambito di applicazione della direttiva

(UE) 2019/1024, che esclude i dati che non sono accessibili per motivi di riservatezza commerciale o statistica e i dati che figurano in opere o in altro materiale su cui terzi detengono diritti di proprietà intellettuale [...]»; «12. [...] I dati detenuti da istituti culturali, quali biblioteche, archivi e musei, nonché orchestre, compagnie d'opera o di balletto e teatri, e da istituti di istruzione non dovrebbero rientrare nel campo di applicazione del presente regolamento in quanto le opere e gli altri documenti in loro possesso sono prevalentemente coperti da diritti di proprietà intellettuale di terzi [...]»; «20. È inoltre di estrema importanza, al fine di preservare una concorrenza leale e l'economia di mercato aperta, salvaguardare i dati protetti di natura non personale, in particolare i segreti commerciali, ma anche i dati non personali che costituiscono un contenuto protetto da diritti di proprietà intellettuale da un accesso illecito che possa portare al furto della proprietà intellettuale o allo spionaggio industriale [...]»). Lo stesso non può dirsi a proposito dei considerando 6 e 18 («6. [...] talune categorie di dati [...] quali dati commerciali riservati, dati soggetti a segreto statistico e dati protetti da diritti di proprietà intellettuale di terzi, compresi segreti commerciali e dati personali, spesso non sono messe a disposizione, [...]»; «18. [...] I dati protetti da diritti di proprietà intellettuale, come pure i segreti commerciali, dovrebbero unicamente essere trasmessi a terzi qualora tale trasmissione sia lecita in virtù del diritto dell'Unione o nazionale o con il consenso del titolare dei diritti [...]»). Tuttavia, e naturalmente, la distinzione, che si ricava dal sistema e dalle categorie del diritto della proprietà intellettuale, ed è trasposta con nettezza sul piano linguistico nei considerando 10, 12 e 20 deve riconoscersi anche ove essa non è riflessa con chiarezza nella lettera della legge, come avviene nei considerando 6 e 18.

³²L'art. 2 par. 1 n. 20 del Regolamento *Data Governance Act* (UE) 2022/868 prevede la seguente definizione di ambiente di trattamento (di dati) sicuro «ambiente di trattamento sicuro»: l'ambiente fisico o virtuale e i mezzi organizzativi per garantire la conformità al diritto dell'Unione, quale il regolamento (UE) 2016/679, in particolare per quanto riguarda i diritti degli interessati, i diritti di proprietà intellettuale e la riservatezza commerciale e statistica, l'integrità e l'accessibilità, per garantire il rispetto del diritto dell'Unione e nazionale applicabile, e per consentire all'entità che fornisce l'ambiente di trattamento sicuro di determinare e controllare tutte le azioni di trattamento dei dati, compresi la visualizzazione, la conservazione, lo scaricamento, l'esportazione dei dati e il calcolo dei dati derivati mediante algoritmi computazionali».

³³L'art. 3, par. 1 del Regolamento *Data Governance Act* (UE) 2022/868 non distingue letteralmente tra i dati e i beni immateriali che formano oggetto di diritti di proprietà intellettuale, ed anzi adotta nella prima proposizione un *wording* che potrebbe far pensare ad una diretta tutela dei dati, ciò che è però da escludere radicalmente leggendo tale proposizione in combinato disposto con la lettera c) dove è fatto chiaro che la protezione può essere solo indiretta ossia solo in quanto discendente ovvero conseguente nei fatti alla protezione dei diritti di proprietà intellettuale che hanno il loro proprio oggetto consistente in beni immateriali e non nei "dati" in sé considerati: «Articolo 3 - *Categorie di dati* - 1. Il presente capo si applica ai dati detenuti da enti pubblici, che sono protetti per motivi di: a) riservatezza commerciale, compresi i segreti commerciali, professionali o d'impresa; b) riservatezza statistica; c) protezione dei diritti di proprietà intellettuale di terzi; o d) protezione dei dati personali, nella misura in cui tali dati non rientrano nell'ambito di applicazione della direttiva (UE) 2019/1024».

³⁴L'art. 5, par. 3 lett. a), punto (ii) del Regolamento *Data Governance Act* (UE) 2022/868 chiarisce che la protezione dei diritti di proprietà intellettuale attiene ai "contenuti" (*content*) non ai "dati": «3. Gli enti pubblici garantiscono, conformemente al diritto dell'Unione e nazionale, la tutela della natura protet-



ta dei dati. Essi garantiscono il rispetto dei requisiti seguenti: a) concedere l'accesso per il riutilizzo dei dati soltanto qualora l'ente pubblico o l'organismo competente abbia garantito, in seguito alla richiesta di riutilizzo, che i dati sono stati: i) anonimizzati, nel caso di dati personali; e ii) modificati, aggregati o trattati mediante qualsiasi altro metodo di controllo della divulgazione, nel caso di informazioni commerciali riservate, compresi i segreti commerciali o *i contenuti protetti da diritti di proprietà intellettuale*;». Il paragrafo 7 del medesimo articolo 5 così prevede: «7. Il riutilizzo dei dati è consentito solo nel rispetto dei diritti di proprietà intellettuale. Il diritto del costituente di una banca di dati di cui all'articolo 7, paragrafo 1, della direttiva 96/9/CE non è esercitato dagli enti pubblici al fine di impedire il riutilizzo dei dati o di limitarlo oltre i limiti stabiliti dal presente regolamento». Nel paragrafo 10 del medesimo articolo 5, la lettera della norma non riflette la distinzione tra dati e beni immateriali che formano oggetto dei diritti di proprietà intellettuale (distinzione che deve comunque riconoscersi, come già osservato): «10. Gli enti pubblici trasmettono dati riservati non personali o dati protetti da diritti di proprietà intellettuale a un riutilizzatore che intende trasferire tali dati a un paese terzo diverso da un paese designato in conformità del paragrafo 12 solo se il riutilizzatore si impegna contrattualmente: [...]». Il paragrafo 12 del medesimo articolo 5 così prevede: «12. Se giustificato da un numero considerevole di richieste in tutta l'Unione riguardanti il riutilizzo di dati non personali in determinati paesi terzi, la Commissione può adottare atti di esecuzione in cui si dichiara che le disposizioni legislative, di vigilanza e in materia di applicazione di un paese terzo: a) garantiscono una protezione della proprietà intellettuale e dei segreti commerciali sostanzialmente equivalente a quella garantita dal diritto dell'Unione; [...]».

³⁵Alle "informazioni contenute nei dati" e ai "contenuti protetti da diritti di proprietà intellettuale" fa riferimento l'art. 7, par. 4 lett. c) del Regolamento *Data Governance Act* (UE) 2022/868 a proposito del sostegno che deve essere fornito dagli "organismi competenti" ivi previsti: «c) fornire assistenza tecnica per la pseudonimizzazione e garantire il trattamento dei dati in modo da tutelare efficacemente la vita privata, la riservatezza, l'integrità e l'accessibilità delle *informazioni contenute nei dati* per i quali è consentito il riutilizzo, comprese le tecniche di anonimizzazione, generalizzazione, soppressione e randomizzazione dei dati personali o altri metodi all'avanguardia di tutela della vita privata, e la cancellazione di informazioni commerciali riservate, tra cui segreti commerciali o *contenuti protetti da diritti di proprietà intellettuale*».

³⁶L'art. 2 par. 1 n. 11 del Regolamento *Data Governance Act* (UE) 2022/868 prevede la seguente definizione di servizi di intermediazione di dati: «11) "servizio di intermediazione dei dati": un servizio che mira a instaurare, attraverso strumenti tecnici, giuridici o di altro tipo, rapporti commerciali ai fini della condivisione dei dati tra un numero indeterminato di interessati e di titolari dei dati, da un lato, e gli utenti dei dati, dall'altro, anche al fine dell'esercizio dei diritti degli interessati in relazione ai dati personali, *ad esclusione almeno di*: a) [...] b) servizi il cui obiettivo principale è l'intermediazione di contenuti protetti da diritto d'autore; [...]».

³⁷Art. 2, par. 1, n. 1) della Proposta di *Data Act*: «"dati": qualsiasi rappresentazione digitale di atti, fatti o informazioni e qualsiasi raccolta di tali atti, fatti o informazioni, anche sotto forma di registrazione sonora, visiva o audiovisiva».

³⁸Art. 35 della Proposta di *Data Act*: «Articolo 35 - *Banche dati che contengono determinati dati* – Il diritto "sui generis" di cui all'articolo 7 della direttiva 96/9/CE non si applica alle banche dati che contengono dati ottenuti o generati dall'uso di un prodotto o di un servizio correlato per non ostacolare l'esercizio del diritto degli utenti di accedere a tali dati e utilizzarli conformemente all'articolo 4 del presente regolamento

o di condividere tali dati con terzi conformemente all'articolo 5 del presente regolamento».

³⁹Art. 2, par. 1, n. 2) della Proposta di *Data Act*: «"prodotto": un bene materiale e mobile, anche quando incorporato in un bene immobile, che ottiene, genera o raccoglie dati relativi al suo utilizzo o al suo ambiente e che è in grado di comunicare dati tramite un servizio di comunicazione elettronica accessibile al pubblico e la cui funzione primaria non è la conservazione e il trattamento dei dati».

⁴⁰Art. 2, par. 1, n. 3) della Proposta di *Data Act*: «"servizio correlato": un servizio digitale, anche software, incorporato in un prodotto o interconnesso con esso in modo tale che la sua assenza impedirebbe al prodotto di svolgere una delle sue funzioni».

⁴¹Il considerando 17 della Proposta di *Data Act* specifica che i dati generati dai prodotti e dai servizi correlati che costituiscono l'oggetto della normativa «non riguardano i dati risultanti da un'elaborazione software che calcoli dati derivati da tali dati, in quanto *tale elaborazione software può essere soggetta a diritti di proprietà intellettuale*».

⁴²Il considerando 28 della Proposta di *Data Act* prevede *inter alia* che «[...] *Nel trattamento dei dati dovrebbero essere rispettati eventuali segreti commerciali o diritti di proprietà intellettuale*. [...]».

⁴³Il considerando 77 della Proposta di *Data Act* richiama la protezione dei diritti di proprietà intellettuale a proposito delle cautele per il trasferimento di dati non personali in paesi terzi: «In altri casi possono verificarsi situazioni in cui una richiesta di trasferimento di dati non personali o di accesso agli stessi basata sulla legislazione di un paese terzo sia in conflitto con l'obbligo di proteggere tali dati a norma della legislazione dell'Unione o nazionale, in particolare per quanto riguarda la tutela dei diritti fondamentali dell'individuo, quali il diritto alla sicurezza e il diritto a un ricorso effettivo, o gli interessi fondamentali di uno Stato membro connessi alla sicurezza nazionale o alla difesa, nonché la protezione dei dati commercialmente sensibili, inclusa la protezione dei segreti commerciali, e *la protezione dei diritti di proprietà intellettuale* [...] Ove possibile in base alle condizioni della richiesta di accesso ai dati dell'autorità del paese terzo, è opportuno che il fornitore di servizi di trattamento dei dati sia in grado di informare il cliente i cui dati sono richiesti al fine di verificare la presenza di un potenziale conflitto tra tale accesso e le norme dell'Unione o nazionali, come quelle sulla protezione dei dati commercialmente sensibili, compresa la protezione dei segreti commerciali, dei *diritti di proprietà intellettuale* e degli impegni contrattuali in materia di riservatezza».

⁴⁴La Proposta di *Data Act* contiene regole armonizzate *inter alia* (i) sulla messa a disposizione di dati generati dall'uso di un prodotto o di un servizio correlato (come ivi definiti) all'utente di tale prodotto o servizio; (ii) sulla messa a disposizione di dati da parte dei *data holders* ai *data recipients* (come rispettivamente ivi definiti); (iii) sulla messa a disposizione di dati da parte dei *data holders* a organi del settore pubblico o istituzioni dell'Unione, agenzie o organi, laddove si verifichi una necessità eccezionale, per l'esecuzione di un compito svolto nel pubblico interesse.

⁴⁵Art. 70-ter, co. 1, l.a.: «Sono consentite le riproduzioni compiute da organismi di ricerca e da istituti di tutela del patrimonio culturale, per scopi di ricerca scientifica, ai fini dell'estrazione di testo e di dati da opere o da altri materiali disponibili in reti o banche di dati cui essi hanno lecitamente accesso, *nonché la comunicazione al pubblico degli esiti della ricerca ove espressi in nuove opere originali*».

⁴⁶Nell'impianto dell'art. 4 CDSMD, l'autorizzazione da parte dei titolari dei diritti si reputa fornita *ex lege* in assenza di un loro esplicito diniego o "riserva", c.d. *opt-out*, e sempre che si tratti di materiali ai quali i terzi hanno "legalmente accesso".



Nella recente consultazione pubblica promossa dal Governo del Regno Unito (v. *supra* nel testo) lo schema di tale soluzione era tra le quattro opzioni considerate (opzione 3). Il Governo del Regno Unito ha annunciato che procederà invece nella direzione dell'eliminazione di qualunque restrizione all'eccezione di TDM (opzione 4).

⁴⁷Sulla tecnica del c.d. *crawling*, cfr., tra gli altri, le spiegazioni e le posizioni di: C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and Data Mining: articles 3 and 4 of the directive 2019/790/EU*, cit., p. 7; A. OTTOLIA, *Big Data*, cit., p. 35 ss.; M.L. MONTAGNANI, G. AIME, *Il text and data mining*, cit., p. 382; J.P. TRIAILLE et al., *Study on the Legal Framework*, cit., p. 10; R. DUCATO, A. STROWEL, *Ensuring Text and Data Mining*, cit., p. 13 ss.

⁴⁸Può essere interessante segnalare l'analisi sulla relazione tra *machine legibility* e TDM condotta nella diversa prospettiva della protezione del consumatore da R. DUCATO, A. STROWEL, *Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to "Machine Legibility"*, in "International Review of Intellectual Property and Competition Law", vol. 50, 2019, n. 6, p. 649 ss.

⁴⁹Sulle radici storica e filosofica dei problemi posti dalle tecnologie digitali alla concezione della proprietà intellettuale, cfr. R. POZZO, *Immanuel Kant on intellectual property*, in "Trans/Form/Ação", vol. 29, 2006, n. 2, p. 11 ss.

⁵⁰Cfr. P. GUARDA, *Il regime giuridico dei dati*, cit., p. 100 ss., spec. p. 104; A. STROWEL, *Reconstructing the reproduction and the communication*, cit., p. 203 ss.; ID., *Big data and data appropriation in the EU*, in T. Aplin (ed.), "Research Handbook on Intellectual Property and Digital Technologies", 2018, p. 119.

⁵¹CGUE 29 luglio 2019, *Pelham GmbH e altri c. Ralf Hütter e Florian Schneider-Esleben*, C-476/1, citato *supra* (caso nel quale la Corte di Giustizia ha negato che ricorresse una riproduzione nel caso in cui un musicista, esercitando la propria libertà delle arti, aveva prelevato un campione sonoro da un

fonogramma altrui al fine di integrarlo, in forma modificata e non riconoscibile all'ascolto umano, in un nuovo fonogramma).

⁵²Per una ricerca delle nozioni giuridiche di informazioni e dati, cfr. P. GUARDA, *Il regime giuridico dei dati*, cit., p. 11 ss.; A. OTTOLIA, *Big Data*, cit. p. 14 ss.; L.A. BYGRAVE, *Information Concepts in Law: Generic Dreams and Definitional Daylight*, in "Oxford Journal of Legal Studies", vol. 35, 2015, n. 1, p. 91 ss.; A. STROWEL, *Big Data and Data Appropriation in the EU*, in T. Alpin, Cheltenham (eds.), "Research Handbook on Intellectual Property and Digital Technologies", Edward Elgar, 2020, p. 107; H. ZECH, *Data as a tradeable commodity. Implications for Contract Law*, in J. Drexel (ed.), "Proceedings of the 18th EIPIN Congress: The New Data Economy between Data Ownership, Privacy and Safeguarding Competition", Edward Elgar Publishing, 2017; H. ZECH, *Data as a tradeable commodity*, in A. De Franceschi (ed.), "European Contract Law and the Digital Single Market. The Implications of the Digital Revolution", Intersentia, 2016, p. 53 ss. Non limitatamente all'ambito di rilevanza giuridica, cfr. R. KITCHIN, *A Critical Analysis of Big Data, Open Data and Data Infrastructures*, 2 ed., Sage Publishing, 2021.

⁵³V. ad esempio artt. 4, 5 e 6 della Proposta di *Data Act* COM(2022) 68, del 23 febbraio 2022 dove si prevedono espressi divieti di utilizzare i dati per certe finalità. In particolare si vieta allo *user* e al *data recipient* di usare i dati generati dall'utilizzo di un prodotto o di un servizio correlato (come definiti nella proposta di *Data Act*) per sviluppare un prodotto concorrente con il prodotto da cui provengono i dati; e si vieta al *data holder* di usare i dati generati da un prodotto o da un servizio correlato (come definiti nella proposta di *Data Act*) per pregiudicare la posizione commerciale dello *user* o del *data recipient*.

⁵⁴Per la cui dimostrazione si rinvia a M. SENFTLEBEN, *Compliance of National TDM Rules with International Copyright Law*, cit., spec. p. 7 ss. e p. 23 ss.

⁵⁵V. *retro sub* nota 22.

* * *

The inexistence of an exclusive right of Text and Data Mining (TDM)

Abstract: The essay analyzes the provisions of the Copyright in the Digital Single Market Directive (Directive (EU) 2019/790: CDSMD) relevant to the so-called Text and Data Mining exception, in order to establish whether the need for an exception or a license as contemplated by Articles 3 and 4 CDSMD is always implicated for the carrying out of the 'Text and Data Mining' operations. On the basis of an interpretation aligned with the most recent case law of the European Court of Justice on the notion of reproduction (in particular, the *Pelham* decision), the negative conclusion is illustrated, and the thesis is presented that the reproduction of machine readable data shall not entail a reserved activity of reproduction or extraction as meant by the legal provisions of Articles 3 and 4 CDSMD. The author also argues that the automated analytical technique aimed at analysing text and data in digital form (i.e. the TDM activity according to the relevant definition), having ordinarily as an object machine readable data only, should not give rise in practice under ordinary circumstances to a problem of compliance with the copyright norms considered by Articles 3 and 4 CDSMD. The author also puts forward the argument that under EU law "data" do not form the subject matter neither of the rights of reproduction and extraction contemplated in articles 3 and 4 CDSMD nor of any other IP rights. On this basis, the article concludes by affirming the inexistence of an exclusive right of TDM.

Keywords: Copyright in the Digital Single Market Directive (CDSMD) – Text and Data Mining (TDM) – Artificial Intelligence – Right of reproduction – *Sui generis* right – Right of extraction