

Via Legnano, 5
26900 Lodi – Italy

tel +39 0371 420 001
fax +39 0371 420 001

simone@aliprandi.org
simone@pec.aliprandi.org

<https://aliprandi.org/>
<https://array.eu>

VAT No. /P. IVA 05871760962

Lodi, 27 febbraio 2020

Spett.
Arpae - Agenzia regionale per la
prevenzione, l'ambiente e l'energia
dell'Emilia-Romagna
Via Po, 5 – 40139 Bologna

via email

Progetto MISTRAL (Consulenza legale) Report n. 4 – Corretta gestione di alcuni scenari concreti e possibili soluzioni

Scopo del presente report è quello di fornire indicazioni operative sui più probabili scenari concreti che il Progetto Mistral si troverà a dover fronteggiare in materia di copyright e licensing dei dati. Questo report si fonda sulle argomentazioni di inquadramento teorico già espresse nei report precedenti e sui casi d'uso emersi nei vari incontri con i partner del progetto tenutisi negli scorsi mesi (e dei quali è stata tenuta traccia in apposite minute).

L'importante distinzione tra costituente di dataset e mero collettore di dataset preesistenti

Prima di entrare nel merito dei vari scenari sottoposti alla mia attenzione, è il caso di chiarire un aspetto fondamentale che poi si riflette su tutto il meccanismo di rilascio e di licenziamento dei dataset. Infatti, a seconda che il Progetto Mistral si ponga come costituente di dataset oppure mero collettore di dataset preesistenti, cambiano le prospettive e le responsabilità anche sul piano della gestione del copyright e delle licenze. Indubbiamente, nel caso in cui il portale Mistral si limiti a raccogliere e ripubblicare dataset provenienti da altre fonti, le responsabilità dei partner del progetto saranno minori e si fermeranno al dover verificare i metadati di licenza sui singoli dataset e ad esporre correttamente tali metadati anche sul portale.

Nel momento in cui il portale Mistral si spinga invece a svolgere attività di validazione dei dati, riorganizzazione dei dati al fine di renderli meglio fruibili dagli utenti, estrazione e reimpiego dei dati dai dataset input al fine di fornire rielaborazioni numeriche o grafiche, la responsabilità dei partner del progetto sarebbe maggiore poiché assumerebbero un ruolo attivo di licenziante dei dataset prodotti come frutto di estrazione e reimpiego. In questo secondo scenario quindi si rende necessaria una verifica sulla compatibilità delle varie licenze dei dataset input e sulla correttezza delle licenze applicate "ex novo" ai dataset prodotti ed esposti dal portale.

Caso 1: aggregazione di dataset preesistenti

SCENARIO 1.1: Tizio realizza il Dataset A con le temperature registrate nel 2019 in tutti i comuni della Provincia Autonoma di Bolzano; Caio realizza un autonomo Dataset B con le temperature registrate nel 2019 nella Provincia Autonoma di Trento. Sempronio vuole realizzare un Dataset

Other Array Members:

Avv. Carlo Piana
Ordine di Milano

Avv. Alberto Pianon
Ordine di Vicenza

Avv. Francesco Paolo Micozzi
Ordine di Cagliari

Avv. Giovanni Battista Gallus
Ordine di Cagliari

Avv. Guglielmo Troiano
Ordine di Milano

C con tutte le temperature registrate nel 2019 nella Regione Trentino-Alto Adige e dunque vuole aggregare Dataset A e Dataset B in un più ampio Dataset C.

QUESITO 1.1: Dataset C dev'essere considerato come un dataset derivato da Dataset A e da Dataset B? E Dataset C deve sottostare comunque alle condizioni delle licenze dei Dataset A e B?

RISPOSTA al quesito 1.1: Sì, Dataset C è un dataset derivato da Dataset A e da Dataset B; di conseguenza deve rispettare sia le condizioni della licenza di Dataset A, sia le condizioni della licenza di Dataset B.

SCENARIO 1.2: Tizio realizza il Dataset A con le temperature registrate nel 2019 in tutti i comuni della Provincia Autonoma di Bolzano; Caio realizza un autonomo Dataset B con le temperature registrate nel 2019 nella Provincia Autonoma di Trento. Sempronio vuole fornire i dati relativi a tutte le temperature registrate nel 2019 nella Regione Trentino-Alto Adige e dunque decide di diffondere Dataset A e Dataset B in un unico file di archiviazione (tipo ZIP), ma mantenendoli tecnicamente separati e indicando per ciascuno la titolarità del copyright e la licenza applicata.

QUESITO 1.2: Il file ZIP dev'essere considerato come un dataset derivato da Dataset A e da Dataset B? Come gestire la questione licenze in questo caso?

RISPOSTA al quesito 1.2: No, il file ZIP non costituisce un dataset a sé; Dataset A e Dataset B rimangono separati e dunque non si viene a creare un rapporto di derivazione. La questione licenze si risolve indicando con chiarezza nei credits di ciascun dataset le informazioni sul titolare dei diritti e sulla licenza applicata.

Caso 2: semplice rielaborazione matematica di dati provenienti da dataset preesistenti

SCENARIO 2: Tizio realizza il Dataset A con le temperature registrate in tutti i comuni d'Italia ora per ora (quindi 24 valori per ciascun comune). Caio utilizza i dati contenuti nel Dataset A per realizzare un Dataset B più semplificato che non fa altro che dare un valore medio giornaliero per ciascun comune (quindi da 24 valori per comune si passa a 1 valore per comune). Questo valore è ottenuto senza un'apprezzabile attività intellettuale bensì attraverso è un banale calcolo matematico (media). Il Dataset B non contiene una estrazione vera e propria dei valori del Dataset A perché nel Dataset B non si trovano valori presenti nel Dataset A; e dal Dataset B non è possibile risalire ai valori più dettagliati del Dataset A.

QUESITO 2: Dataset B dev'essere considerato come dataset derivato da Dataset A?

RISPOSTA al quesito 2: Sì. L'attività di rielaborazione matematica, anche se è banale e non richiede un intervento intellettuale/creativo apprezzabile, integra quello che la Direttiva 96/9/CE chiama "reimpiego" (si veda la definizione di cui all'art. 7, co. 2).

Caso 3: mera validazione di dati contenuti in un dataset preesistente

SCENARIO 3.1: Tizio realizza il Dataset A con la pressione atmosferica registrata in un determinato luogo a cadenza giornaliera. Caio analizza i dati presenti in Dataset A per procedere a una loro validazione scientifico-tecnica e pubblica un Dataset A-bis

sostanzialmente identico, ma validato; cioè senza alcun intervento sui dati ma solo con l'indicazione che il dataset è stato da lui validato.

QUESITO 3.1: Dataset A-bis dev'essere considerato come dataset derivato da Dataset A?

RISPOSTA al quesito 3.1: Sì. Dataset A-bis non potrebbe esistere se non fosse derivato da Dataset A. Tra l'altro anche la mera attività di verifica di un database rientra nelle attività rilevanti ai fini della particolare tutela "sui generis" che la Direttiva 96/9/CE riserva alle banche dati non creative. L'art. 7, co. 1, precisa infatti che il costituente di una banca dati non creativa è titolare del diritto sui generis quando il conseguimento, la *verifica* e la presentazione del contenuto di una banca dati attestino un investimento rilevante sotto il profilo qualitativo o quantitativo.

SCENARIO 3.2: Tizio realizza Dataset A con la pressione atmosferica registrata in un determinato luogo a cadenza giornaliera. Caio analizza i dati presenti in Dataset A per procedere a una loro validazione scientifico-tecnica e pubblica un Dataset A-bis che riporta una colonna aggiuntiva con l'indicazione dei dati validati e dei dati non validati.

QUESITO 3.2: Dataset A-bis dev'essere considerato come dataset derivato da Dataset A?

RISPOSTA al quesito 3.2: Sì. Come nello Scenario 3.1, Dataset A-bis non potrebbe esistere se non fosse derivato da Dataset A. Oltre a tenere in considerazione quanto esposto nella risposta al Quesito 3.1, si deve qui aggiungere che anche la semplice aggiunta di una colonna rappresenta un'attività di reimpiego del dataset; attività che, benché possa risultare marginale a livello quantitativo, ha indubbiamente rilevanza sul piano qualitativo poiché va a modificare il concept generale del dataset, il suo valore e la sua possibilità di essere utilizzato da più utenti.

Caso 4: dataset realizzati attraverso un processo di modellazione di dati prelevati da un dataset preesistente

SCENARIO 4: Tizio realizza il Dataset A con le temperature registrate in una determinata zona negli ultimi tre giorni; Caio realizza un autonomo Dataset B con il tasso di umidità registrato nella stessa zona negli ultimi tre giorni; Sempronio realizza un autonomo Dataset C con la pressione atmosferica registrata nella stessa zona negli ultimi tre giorni. Mevio, ha realizzato un modello di calcolo che incrocia i dati presenti nei Dataset A, B e C e permetta di creare un Dataset Z con le previsioni meteo per le prossime 24 ore.

QUESITO 4: Il Dataset Z dev'essere considerato come dataset derivato dai Dataset A, B e C?

RISPOSTA al quesito 4: Sì. Il modello di calcolo può generare i suoi output e andare così a popolare il Dataset Z solo con un'attività di estrazione e reimpiego di parti sostanziali dei Dataset A, B e C; ne consegue che le licenze dei tre dataset di input devono essere tra loro compatibili; il Dataset Z dovrà rispettare le indicazioni di titolarità dei diritti (c.d. *attribution*) di tutti i tre database input; e la licenza scelta per il Dataset Z dovrà comunque essere compatibile con tutte e tre le licenze dei dataset di input.

Caso 5: elaborazione/rappresentazione grafica del contenuto di un dataset

SCENARIO 5.1: Tizio realizza il Dataset A con le precipitazioni registrate negli ultimi 30 giorni in tutta la Regione Emilia-Romagna; Sempronio vuole realizzare una rappresentazione grafica dei dati contenuti in Dataset A in cui si mostra sulla mappa della regione la distribuzione delle precipitazioni attraverso colori, segni grafici e indicazioni numeriche. La mappa è un file d'immagine raster o vettoriale e non permette la ricostruzione precisa del database in ingresso.

QUESITO 5.1: La mappa è da considerare come derivata dal Dataset A e deve sottostare comunque alle condizioni della licenza del Dataset A?

RISPOSTA al quesito 5.1: Per rispondere a questo quesito bisogna innanzitutto precisare che non vi è una definizione univoca di "opera derivata"; addirittura nel nostro ordinamento l'espressione "opera derivata" non è presente in alcuna norma di legge ed è utilizzata più propriamente l'espressione "elaborazione creativa". Per evitare discrasie interpretative, è prassi inserire nelle varie licenze una definizione ad hoc. Vista questa premessa, la risposta al quesito non può che rimandare alle definizioni presenti nelle due principali licenze, cioè CC BY-SA 4.0 e la ODbL. Come già illustrato la definizione di "derivazione" è la principale causa di incompatibilità tra le due licenze. La ODbL infatti fa una importante distinzione tra Derivative Dataset e Produced Work; e il caso della realizzazione di una grafica o di una mappa basata sui dati di un dataset è proprio l'esempio più classico di "Produced Work". Nella definizione fornita dalla licenza un "Produced Work" è "a work (such as an image, audiovisual material, text, or sounds) resulting from using the whole or a Substantial part of the Contents (via a search or other query) from this Database, a Derivative Database, or this Database as part of a Collective Database." Dunque secondo le previsioni della licenza ODbL, una grafica o una mappa basata sui dati di un dataset non è da considerare alla stregua di un Derivative Database e dunque non comporta le restrizioni di "share alike" che invece dovremmo rispettare qualora realizzassimo un dataset derivato.

Al contrario, la licenza CC BY-SA 4.0 non contiene questa distinzione; quindi una grafica o una mappa basata sui dati di un dataset rilasciato con tale licenza è da considerare alla stregua di un'opera derivata e di conseguenza fa scattare l'obbligo di "share alike".

SCENARIO 5.2: Tizio realizza il Dataset A con le precipitazioni registrate negli ultimi 30 giorni in tutta la Regione Emilia-Romagna e lo espone con la Licenza X; Caio realizza il Dataset B con le temperature registrate negli ultimi 30 giorni in tutta la Regione Emilia-Romagna e lo espone con la Licenza Y. Sempronio vuole realizzare una rappresentazione grafica dei dati contenuti sia in Dataset A sia in Dataset B (tra essi combinati e sovrapposti) in cui si mostra sulla mappa della regione la distribuzione delle precipitazioni e delle temperature attraverso colori, segni grafici e indicazioni numeriche. La mappa è un file d'immagine raster o vettoriale e non permette la ricostruzione precisa del database in ingresso.

QUESITO 5.2: La mappa è da considerare come derivata dai Dataset A e B e deve sottostare alle condizioni di entrambe le licenze X e Y?

RISPOSTA al quesito 5.2: Anche qui vale la stessa risposta fornita al quesito 5.1, con la considerazione aggiuntiva che qui è necessario anche preoccuparsi della compatibilità tra le licenze X e Y.

Caso 6: realizzazione di un'opera letteraria (tesi di dottorato, articolo scientifico, etc.) basata sui dati contenuti in dataset preesistenti

SCENARIO 5.1: Sul portale Mistral vengono esposti due diversi dataset: il Dataset A con i dati delle stazioni a terra di MeteoNetwork rilasciato con licenza ODbL e il Dataset B con i dati delle precipitazioni della rete radar nazionale e distribuiti con licenza CC BY-SA. Tizio vuole realizzare un'opera letteraria di carattere tecnico-scientifico (tesi di dottorato, un articolo su rivista, un report di ricerca) in cui vengono confrontati questi dati, ad esempio calcolando la statistica della differenza tra precipitazione "vista" da radar e la precipitazione misurata dalle stazioni di MeteoNetwork.

QUESITO 5.1: L'opera letteraria dovrebbe sottostare comunque alle condizioni delle licenze dei dataset e dovrebbe quindi preoccuparsi del problema dell'incompatibilità tra le due licenze?

RISPOSTA al quesito 5.1: Questo è un quesito che richiederebbe una vasta gamma di considerazioni sul funzionamento del diritto d'autore. In sintesi, innanzitutto è abbastanza forzato sostenere che un'opera letteraria che commenta dei dati contenuti in una banca dati sia da considerare come un'opera derivata della banca dati stessa; a meno che si tratti di un'opera letteraria interamente incentrata sul commento, sulla rappresentazione grafica, sulla rielaborazione di una parte sostanziale di un'unica banca dati, in modo da rendere la fruizione della banca dati quasi secondaria dato che buona parte dei dati sono già stati ripubblicati nell'opera letteraria (ma appunto si tratta di una situazione limite e non così probabile). Inoltre l'attività di scrittura di opere letterarie di carattere scientifico è coperta da più ampi spiragli di "libera utilizzazione", come ad esempio il diritto di citare e riprodurre brani o parti di opere per uso di critica o di discussione nei limiti giustificati da tali fini e purché non costituiscano concorrenza all'utilizzazione economica dell'opera (art. 70 Legge 633/1941). Ricordiamo comunque che, qualora la ripubblicazione di grafiche e immagini contenenti i dati avvenisse attraverso la rete internet, a titolo gratuito, a bassa risoluzione, per uso didattico o scientifico e senza scopo di lucro, si rientrerebbe nell'eccezione al diritto d'autore di cui al comma 1-bis della stessa norma. Va da sé che ogni riproduzione o citazione dovrà essere accompagnata dalla menzione del titolo dell'opera, dei nomi dell'autore, dell'editore e, se si tratti di traduzione, del traduttore, qualora tali indicazioni figurino sull'opera riprodotta.

Indicazioni sulla modalità di comunicazione agli utenti delle licenze applicate ai dati

Mi viene chiesto di fornire indicazioni pratiche su come è opportuno comunicare agli utenti le varie licenze applicate ai dati a seconda dei vari scenari concreti.

Da un punto di vista strettamente giuridico è sufficiente che il licenziante metta il licenziatario nella condizione di conoscere con chiarezza e senza possibilità di equivoci il regime di copyright dell'opera licenziata.

Come su altri piani, anche qui i problemi maggiori si incontrano con i dataset rilasciati con licenze tipo "share-alike" (CC BY-SA 4.0, ODbL e simili), le quali, per le ragioni di incompatibilità che abbiamo esposto più volte, impongono che i dataset vengano tenuti separati e dunque che anche le informazioni sul regime dei diritti siano ben distinte da dataset a dataset. In caso di dataset rilasciati sotto ODbL si tenga presente la già illustrata distinzione tra "Derivative database" e "Produced Work"; l'obbligo di condividere con la stessa licenza scatta solo in caso di realizzazione di un "Derivative database", mentre la realizzazione di un "Produced work" richiede solo il rispetto della cosiddetta attribuzione.

In caso di dataset rilasciati con licenze di “mera attribuzione” (CC BY 4.0, IODL 2.0 e simili) la situazione è meno complicata, dato che è sufficiente rendere noto chi sia il titolare dei diritti secondo quanto indicato dal testo della licenza.

La clausola “Attribuzione” (art. 3) delle licenze CC in versione 4.0 ha il seguente testo:

Se Tu Condividi il Materiale Concesso in Licenza (anche in forma modificata), Tu sei tenuto a:

- riportare, se fornito dal Licenziante assieme al Materiale Concesso in Licenza, ciò che segue:

- l'identificazione del creatore o dei creatori del Materiale Concesso in Licenza e delle terze parti designate a ricevere l'attribuzione, in qualsiasi maniera ragionevole che sia richiesta dal Licenziante (incluso lo pseudonimo, se designato);*
- l'informativa sul diritto d'autore;*
- l'informativa che si riferisce alla presente Licenza Pubblica;*
- l'informativa contenente esclusioni o limitazioni di responsabilità;*
- l'Uniform Resource Identifier (URI) o il collegamento ipertestuale alla presente Licenza Pubblica nella misura in cui sia ragionevolmente possibile;*

- indicare se Tu hai modificato il Materiale Concesso in Licenza e, nel caso, conservare un'indicazione di ogni modifica precedente; e

- indicare che il Materiale Concesso in Licenza è rilasciato secondo i termini e le condizioni della presente Licenza Pubblica, e includere il testo della, oppure l'URI o il collegamento ipertestuale alla, presente Licenza Pubblica.

Tu puoi adempiere alle condizioni dell'Art. 3(a)(1) in qualsiasi maniera ragionevole, rispetto al mezzo di comunicazione, al supporto, agli strumenti e al contesto all'interno del quale Tu Condividi il Materiale Concesso in Licenza. Ad esempio, può essere ragionevole soddisfare le condizioni fornendo l'URI o il collegamento ipertestuale a una risorsa che includa le informazioni richieste.

Su richiesta del Licenziante, nella misura in cui ciò sia ragionevolmente praticabile, Tu devi rimuovere ognuna delle informazioni richieste dall'Art. 3(a)(1) (A).

Se Tu Condividi Materiale Elaborato da Te prodotto, la Licenza dell'Elaboratore da Te applicata non deve impedire ai destinatari del Materiale Elaborato di adempiere ai termini e alle condizioni della presente Licenza Pubblica.

La clausola “Notice for using output (Contents)” (Art. 4.3) della licenza ODbL ha il seguente testo:

Creating and Using a Produced Work does not require the notice in Section 4.2. However, if you Publicly Use a Produced Work, You must include a notice associated with the Produced Work reasonably calculated to make any Person that uses, views, accesses, interacts with, or is otherwise exposed to the Produced Work aware that Content was obtained from the Database, Derivative Database, or the Database as part of a Collective Database, and that it is available under this License.

- Example notice. The following text will satisfy notice under Section 4.3:

[Contains information from DATABASE NAME, which is made available here under the Open Database License (ODbL)]

DATABASE NAME should be replaced with the name of the Database and a hyperlink to the URI of the Database. "Open Database License" should contain a hyperlink to the URI of the text of this License. If hyperlinks are not possible, You should include the plain text of the required URI's with the above notice.

Ovviamente non si pongono particolari problemi in caso di dataset rilasciati con public domain waiver (CC0, PDDL): è possibile infatti evitare anche di citare la fonte; la citazione della fonte può essere mantenuta in via facoltativa per una questione di tracciabilità delle fonti e di autorevolezza del dato.

Per tutto il resto è opportuno fare riferimento alle Linee Guida sull'OpenData dell'AgID (in particolare il Capitolo 8 – Modello per i metadati) e in generale alle best practice in materia di metadato delle banche dati. Infatti le informazioni relative al licenziante e al tipo di licenza applicata sono normalmente espresse in forma di metadati "machine readable".

Come comportarsi in caso di cambio licenza sui dati in entrata

I dataset input per il Portale Mistral saranno dinamici e prelevati costantemente online con un'attività di scraping/parsing automatizzata; e le licenze di alcuni dataset input potrebbero cambiare nel corso del progetto. Anche qualora il Portale Mistral si ponesse come mero ripubblicatore/riaggregatore di dataset preesistenti, è necessario che vi sia consistenza tra le licenze esposte sul portale e le licenze esposte sui siti web delle fonti originali. Tra l'altro, in caso di cambio di licenza che porti verso una licenza più restrittiva rispetto a quella precedente, potrebbe rendersi necessario escludere tale dataset da alcune attività e da alcune interazioni con altri dataset (pensiamo al classico caso di un cambio licenza verso una licenza CC BY-SA 4.0 che renderebbe il dataset non più compatibile con altri).

Dal momento che le fonti di dataset input sono molte ed è oggettivamente impossibile che vi sia un controllo costante di tipo "umano" costante, giorno per giorno, su ciascuna fonte, consiglio di predisporre uno script che, rilevando e monitorando il metadato "licenza", sia in grado di segnalare e inviare una notifica in caso di modifica delle licenze.

(Avv. Simone Aliprandi)

