

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, 18 novembre 2019

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. 2 – Proprietà intellettuale e licenze: inquadramento generale

Scopo del presente report è quello di fornire un quadro il più completo possibile delle questioni emergenti in tema di gestione delle licenze d'uso in entrata e in uscita sia sul software sia sui dati.

Questo report è realizzato sulla base della documentazione sottopostami in fase di studio preliminare (a tal proposito si veda anche il Report n. 1) e sulla base di quanto emerso negli incontri con i responsabili del progetto e con i partner tenuti fino al 10 novembre 2019.

### 1. Gestione delle licenze software

#### 1.a. Il licensing di software con approccio open: scenario teorico

Rispetto all'ambito delle banche dati (di cui parleremo più avanti), quello del software risulta meno problematico. Le licenze open per il software sono un fenomeno che ha una storia ben più radicata negli anni (i primi esperimenti risalgono infatti alla fine degli anni 80); possiamo quindi disporre di tutta una serie di indicazioni e *best practice* nonché di qualche precedente giurisprudenziale a cui fare riferimento. Nonostante le licenze ad oggi considerate *compliant* con la Open Source Definition siano circa un'ottantina (si veda l'elenco completo e aggiornato sul sito <https://opensource.org/licenses/>), tutte sono grossomodo riconducibili alle tre macrocategorie elaborate dalla letteratura di settore: licenze permissive, licenze con copyleft forte, licenze con copyleft debole.

Le prime (licenze permissive) sono licenze molto semplici che consentono un utilizzo molto ampio del software e richiedono come unica condizione quella della cosiddetta "attribuzione"; tuttavia non pongono alcuna condizione affinché il codice sorgente rimanga aperto anche su eventuali opere derivate. Esempi classici di queste licenze si trovano nei progetti software nati in ambito accademico (si vedano la licenza BSD e la licenza MIT).

Le licenze copyleft sono quelle che impongono che anche le opere derivate vengano distribuite sotto la stessa licenza dell'opera originale. A seconda dell'ampiezza di questa condizione, si ha un copyleft "forte" o un copyleft "debole".

L'archetipo della licenza con copyleft forte è la GNU GPL, la licenza principale del progetto GNU e in generale la più popolare delle licenze di software libero, arrivata oggi alla versione 3. Secondo questa licenza, tutte le opere derivate, nella più ampia concezione del termine concessa dal diritto d'autore, devono essere rilasciate sotto la stessa licenza. Esiste poi una

#### 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

variante della Licenza GNU GPL v.3 chiamata Affero GPL o AGPL. Essa è simile alla capostipite GNU General Public License, eccetto per il fatto che ha una sezione aggiuntiva (la numero 13) che si riferisce all'utilizzo del software su una rete di calcolatori. Questa sezione richiede che il codice sorgente, se modificato, sia reso disponibile a chiunque utilizzi l'opera sulla rete. Il codice da fornire non sarà solo quello coperto da AGPL, ma anche tutti i moduli da esso utilizzati, escluse naturalmente le librerie di sistema.

Le licenze "copyleft debole" prevedono che l'obbligo di utilizzare la medesima licenza per le opere derivate sia limitato alle modifiche al programma stesso, ma non si allarga a opere più ampie che ne incorporano il codice nel caso in cui questa parte di codice rimanga individuabile e separata dal resto (come nel caso delle librerie di funzioni). L'archetipo della licenza di copyleft debole è la GNU Lesser General Public License (o LGPL). Tale licenza è stata concepita per alcune librerie software del progetto GNU (infatti l'acronimo originale della LGPL era Library General Public License) in quanto si voleva che tali librerie fossero utilizzabili anche da programmi proprietari; un'altra licenza di copyleft debole di ampio successo è la MPL della Mozilla Foundation.

## 1.b. Software originale e software derivato

Nell'ambito del Progetto MISTRAL è stato realizzato e verrà realizzato del software, che, elaborando i vari dati, implementa i modelli di previsione meteorologica e offre la possibilità di effettuare calcoli specifici e raffigurazioni grafiche di tali dati. Come si deduce già dalla lettura del Grant Agreement e del Consortium Agreement, nonché dalle politiche interne dei singoli partner, tutto il software sviluppato in seno al progetto deve essere rilasciato in modalità open, secondo la definizione più accreditata a livello mondiale di "software open source" rappresentata dall'Open Source Definition della Open Source Initiative (<https://opensource.org/osd>). In effetti l'adozione da parte del Progetto MISTRAL (e in particolare da parte di quei partner del progetto che sono enti pubblici) risulta in linea con le norme giuridiche e le linee guida nazionali in materia (si vedano a tal proposito il "Codice Amministrazione Digitale" e le "Linee guida su acquisizione e riuso software per le pubbliche amministrazioni" rilasciate dall'AgID); dall'altra parte essa garantirà, in prospettiva di futuri sviluppi del progetto, una maggiore possibilità di monitoraggio, manutenzione, aggiornamento del software sviluppato nonché una maggiore probabilità di creazione "spontanea" di una community di sviluppatori che provveda alla sopravvivenza del software.

In questo scenario, è fondamentale tenere ben distinto il software originale, cioè quello sviluppato in seno al progetto dai partner e di cui i partner detengono pieni diritti di utilizzo, da quello realizzato come opera derivata partendo da pacchetti software open source. Per il software originale non ci saranno particolari problemi nella scelta e applicazione della licenza; per il software derivato si dovrà verificare con attenzione la licenza applicata dal titolare dei diritti e, qualora si tratti di una licenza con clausola propagativa, si dovrà comunque mantenere la stessa licenza anche sulle derivazioni.

## 1.c. Le licenze dei pacchetti software originali sviluppati dai partner

Come segnalatomi dai coordinatori del progetto, gli enti partner hanno sviluppato alcuni pacchetti software originali di cui detengono pieni diritti e che hanno deciso di aggiornare e adattare in vista del loro utilizzo in seno al Progetto MISTRAL. I pacchetti sono i seguenti:

- 1) *dballe* (indirizzo GitHub <https://github.com/ARPA-SIMC/dballe/>): software per la gestione dei dati osservativi;
- 2) *arkimet* (indirizzo GitHub <https://github.com/ARPA-SIMC/arkimet/>): software di archiviazione di dati modellistici e osservativi;
- 3) *libsिम* (indirizzo Github <https://github.com/ARPA-SIMC/libsिम/>): software per la postelaborazione dei dati.

Tutti e tre sono rilasciati con la licenza GNU General Public License versione 2.0, licenza che, come spiegato sopra, rappresenta l'archetipo più classico delle licenze con copyleft forte e

che risulta pienamente in linea con le "Linee guida su acquisizione e riuso software per le pubbliche amministrazioni" rilasciate dall'AgID.

### 1.d. Le licenze consigliate

La licenza GNU GPL è indubbiamente una scelta ottimale per i pacchetti software che verranno distribuiti o messi a disposizione del pubblico. A meno che ci siano esigenze particolari in senso contrario, si consiglia comunque un passaggio alla sua versione 3.

Tuttavia negli ultimi anni si è registrato un decremento del software distribuito e messo a disposizione in copie a favore di un incremento del software reso fruibile da remoto su un server internet (cosiddetto "cloud computing"). Dal momento che le clausole della licenza GNU GPL scattano con un atto di distribuzione, in tutti quei casi in cui non vi è una vera distribuzione gli effetti di tale licenza vengono vanificati. Come già spiegato nei paragrafi precedenti, per rispondere a questa esigenza è stata predisposta una variante di questa licenza chiamata GNU Affero GPL. La Free Software Foundation raccomanda questa licenza per ogni software che sia comunemente pensato per applicazioni web e che in generale sia reso disponibile su una rete di calcolatori. Invito quindi i partner del Progetto MISTRAL a considerare questa opzione.

## 2. La gestione delle licenze sui dati

### 2.a. La complessità del diritto sui generis e il concetto di "derivazione" nel mondo delle banche dati

Buona parte dei dati che entrano in gioco nel progetto MISTRAL sono tutelati non tanto da diritto d'autore in senso classico quanto dal cosiddetto diritto *sui generis* del costituente di banche dati istituito dalla Direttiva 96/9/CE. Tale diritto copre operazioni di estrazione e reimpiego di parti sostanziali di banche dati per la cui costituzione è stato necessario un rilevante investimento. Si tratta di un diritto complesso e relativamente nuovo per cui non sempre la giurisprudenza e la dottrina giuridica hanno avuto modo di sviluppare orientamenti consolidati e interpretazioni univoche, sia in merito ai suoi presupposti (il concetto di rilevante investimento, specie quando il costituente è un'istituzione pubblica) sia in merito alla sua estensione (che cosa si intende per "reimpiego"? Che cosa si intende per "parte sostanziale"?).

Inoltre, buona parte dei problemi in tema di *licensing* e riutilizzo di dati deriva dalle clausole che limitano/condizionano l'attività di realizzazione di database "derivati". Il concetto di "opera derivata" proviene infatti dal mondo del diritto d'autore classico (classici esempi di derivazione sono la traduzione di un testo letterario, il remix di un brano musicale, l'adattamento cinematografico di un'opera teatrale); mentre la realizzazione di una banca dati "nuova" partendo da una banca dati preesistente comporta un'attività diversa, che quindi, per essere correttamente inquadrata giuridicamente, necessita di essere definita in modo molto preciso nelle licenze d'uso applicate. Purtroppo non tutte le licenze forniscono una definizione di "database derivato" e le licenze che la forniscono spesso ne forniscono una sostanzialmente diversa da quella di altre licenze pre-esistenti, creando poi seri problemi di compatibilità.

### 2.b. Il licensing di dati e il problema delle licenze che non licenziano

Non tutte le licenze in circolazione licenziano il diritto *sui generis*, cioè non tutte contengono definizioni e clausole che trattano adeguatamente tale diritto, lasciandolo così "non licenziato". Ciò dipende dal fatto che il diritto sui generis è un'invenzione tipicamente europea e le licenze per opere creative più note (come quelle di Creative Commons e della Free Software Foundation) provengono invece dall'ordinamento USA.

Ad esempio, nel caso delle licenze Creative Commons, solo la versione 4.0 ha risolto il problema del diritto sui generis e dunque solo questa versione risulta efficace anche se applicata a banche dati. Oppure bisogna utilizzare licenze proposte da altri progetti e

specificamente pensate con quello spirito: tra queste si annoverano le licenze di Open Data Commons come la ODC-Attribution e la Open Databases License (ODbL).

Un primo step verso il riutilizzo di dataset rilasciati con una licenza pubblica è quindi quello di verificare innanzitutto se tali dataset siano rilasciati con licenze adeguate (che appunto licenziano anche il diritto *sui generis*) e se, in caso contrario, ci sia altra base giuridica per utilizzare liberamente i dati o, in alternativa, se sia necessario contattare i titolari per ottenere specifico permesso.

## 2.c. Tipologia dei dati riutilizzati per il progetto

Mi è stata illustrata la tipologia di dati che entreranno in gioco nel progetto, delineando l'utile distinzione in due macrocategorie: dati derivanti da mera osservazione e dati frutto di modelli previsionali.

Le due tipologie di dati sollevano questioni giuridiche non banali che dovranno essere oggetto di attenta riflessione e di approfondite verifiche. In particolare, per quanto riguarda i dati di mera osservazione, sarà necessario comprendere se essi sono la semplice "fotografia" di una realtà statica presente in natura che chiunque sarebbe in grado di misurare ottenendo lo stesso dato, o se al contrario sussistono diverse variabili che li influenzano e quindi possono portare a misurazioni differenti. Per quanto riguarda invece i dati frutto di modelli saranno necessarie ulteriori accortezze al fine di capire se tali dati debbano essere considerati come strettamente connessi al modello matematico e al software che li producono o se siano invece da considerare come entità a sé stanti.

Nei casi in cui i dati rilasciati dal progetto sono il frutto di un calcolo complesso basato su altre fonti di dati e questa relazione di dipendenza sia stretta ed evidente, cioè in altre parole, nei casi in cui il database Y diffuso come output del progetto non può che essere ottenuto da un calcolo basato sul database X (input), il database Y è da considerare come database derivato da X (con tutte le conseguenze che ciò comporta sul piano delle licenze). Diverso è invece il caso in cui il database Y diffuso come output sia frutto di un'attività di aggregazione e miscelazione che potrebbe anche basarsi su dati input che sono facilmente misurabili da chiunque o comunque sono facilmente disponibili per tutti; in questo caso non vi sarebbe una dipendenza così stretta e l'eventuale dipendenza non sarebbe facilmente dimostrabile. Chiaramente questo approccio non torna utile in tutti quei casi in cui è lo stesso licenziatario a dichiarare espressamente e pubblicamente che i suoi dati output si basano (e quindi riutilizzano) dati tratti da uno specifico licenziante.

## 2.d. Il modello di previsione: che tipo di opera?

Sarà necessaria una particolare riflessione per un corretto inquadramento giuridico dei cosiddetti modelli di previsione. È infatti importante ricordare che i modelli matematici non sono in sé tutelati da alcun istituto del diritto della proprietà intellettuale. Ciò nonostante essi sono comunque tutelabili in via indiretta, ad esempio con copyright nella loro componente software, oppure con segreto industriale quando sono effettivamente mantenuti segreti. Qualora tra gli output del progetto che il consorzio intende rilasciare al pubblico ci fossero anche i modelli di previsione, si dovrà stabilire se considerarli a tutti gli effetti come opere software e dunque applicarvi una licenza d'uso specifica, o se semplicemente considerarli come algoritmo matematico non licenziabile in alcun modo, o ancora se considerarli come un'entità complessa e indissolubile composta di software + algoritmo + banche dati per la quale difficilmente si potrà trovare un'unica licenza.

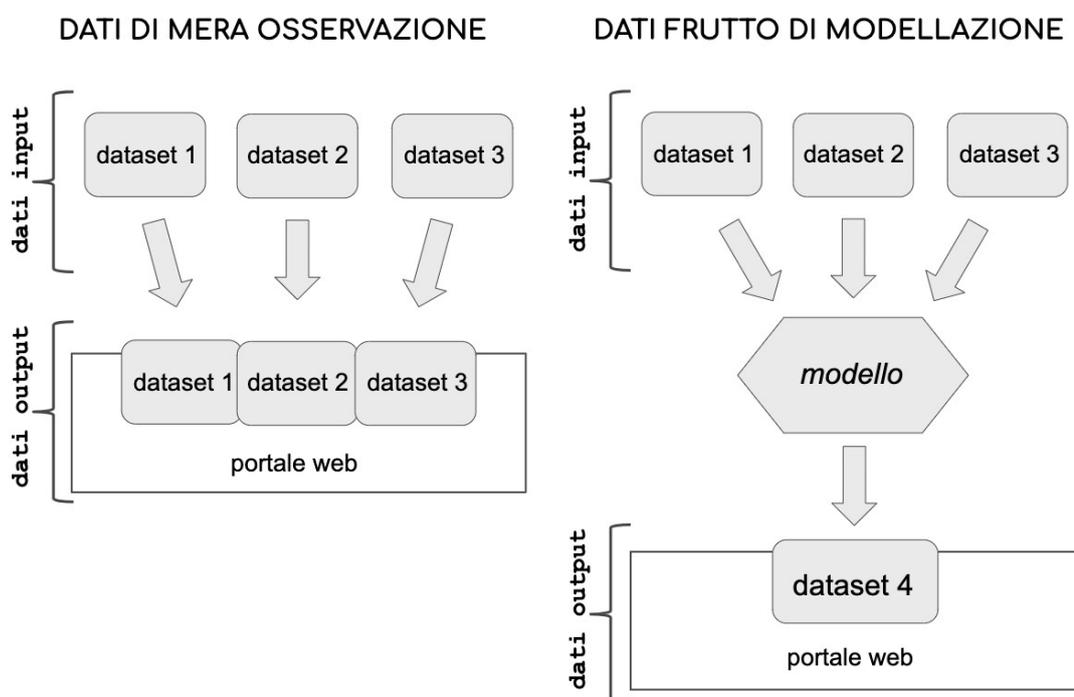
A mio avviso i modelli di previsione, per come mi sono stati illustrati dai partner del progetto, costituiscono un'opera complessa formata da una componente informatica (software e librerie), da una componente matematica (che a seconda dei casi può essere di per sé un'opera dell'ingegno tutelata) e da una componente di banche dati, tutte e tre pensate per funzionare in modo integrato, ma comunque tecnicamente e giuridicamente separabili.

Dal momento che software e banche dati sono tipi di opere dell'ingegno con caratteristiche diverse e sottostanno a un inquadramento giuridico non omogeneo (le banche dati, come detto,

sono tutelate da un diritto diverso rispetto al copyright classico), non è possibile licenziare le due componenti come un "tutt'uno", applicandovi un'unica licenza. Qualora dunque si volessero rilasciare in modalità open anche i modelli di previsione, unica via percorribile sarebbe dunque quella di esporre i modelli esplicitando che la componente software è rilasciata con la licenza X e la componente non software (modello matematico e banca dati) è rilasciata con licenza Y, e consentendo agli utilizzatori di acquisire anche le due componenti separate.

## 2.e. Workflow dei dati in entrata e dei dati in uscita

Sulla base di quanto illustratomi dai partner in merito ai dati in entrata (dati input) e sul tipo di dati che il progetto vorrà produrre e rendere disponibili liberamente al pubblico attraverso il portale web Meteonte (dati output) ho elaborato un diagramma che rappresenta (pur sinteticamente e schematicamente) il workflow dei dati.



Come si può facilmente intuire, il caso dei dati frutto di modellazione è quello che pone maggiori criticità, poiché genera una serie di quesiti relativi alla sussistenza di un rapporto di derivazione tra i dati input e i dati output. Meno problematico è invece il caso dei dati di mera osservazione visto che l'attività del portale in quel caso sarebbe più che altro di raccolta, riorganizzazione, aggregazione e messa a disposizione del pubblico.

## 2.f. Scelta e applicazione delle licenze in uscita: database originali

Come già sottolineato in merito al software, sulle opere interamente originali, cioè create *ex novo* nell'ambito delle attività del progetto, i singoli partner autori o il consorzio potranno scegliere con massima libertà quali licenze applicare tra quelle normalmente consigliate per i cosiddetti opendata. Per la scelta di licenze che siano il più possibile funzionali al concetto di open data e che siano rispettose delle direttive europee in materia di *public sector information*, si rimanda sia alle indicazioni fornite nelle "Linee guida per la valorizzazione del patrimonio informativo pubblico" dell'Agenzia per l'Italia Digital (AgID) sia alle indicazioni dalla Open Knowledge Foundation, ente particolarmente autorevole a livello internazionale su questi temi.

Quest'ultima ha individuato una ristretta cerchia di licenze open data che possono garantire una piena compliance con la Open Definition (vedi <https://opendefinition.org/licenses/>); rielaborando la tabella offerta da OKFN, possiamo arrivare a questo elenco suddiviso in tre categorie:

A) Public domain dedications / no-copyright disclaimers

→ CC0 <https://creativecommons.org/publicdomain/zero/1.0/>

→ PDDL <https://it.okfn.org/pddl/>

B) Attribution licenses

→ CC BY 4.0 <https://creativecommons.org/licenses/by/4.0/>

→ ODC-Attribution <https://opendatacommons.org/licenses/by/index.html>

→ IODL 2.0 <https://www.dati.gov.it/content/italian-open-data-license-v20>

C) Share Alike licenses

→ CC BY 4.0 <https://creativecommons.org/licenses/by-sa/4.0/>

→ ODbL <https://opendatacommons.org/licenses/odbl/summary/index.html>

→ IODL 1.0 [ritirata/deprecata]

Le licenze delle categorie A e B consentono di "liberare" il più possibile i dataset e quindi di aumentarne il loro potenziale di riuso; dunque di certo risultano le opzioni preferibili. Le licenze con clausola propagativa risultano invece problematiche principalmente perché, a valle, pongono problemi di compatibilità con altre licenze. Si noti inoltre che, contrariamente a quanto molti pensano, la clausola propagativa (cosiddetto "share alike") non svolge la stessa funzione che invece la clausola copyleft svolge nel mondo del software (cioè quella di evitare che il codice sorgente venga "chiuso"); perciò non risulta sempre "strategica" ai fini della promozione di grandi community dedite allo sviluppo e alla diffusione di dati aperti.

## 2.g. Gestione delle licenze di database non originali e relativi problemi di compatibilità tra licenze

Va considerato anche il caso di dataset preesistenti utilizzati dal consorzio perché considerati di pubblico dominio o in uno status "open by default" (cioè secondo lo status giuridico definito dall'art. 52 del Codice Amministrazione Digitale); essi non destano particolari preoccupazioni e quindi sono utilizzabili liberamente. Ovviamente è buona prassi verificare che quello sia davvero il loro *status* e che non sussistano altre cause ostative al loro riutilizzo.

Ben più complicato è invece il caso di dati output che consistono in "derivazioni" di banche dati preesistenti utilizzate dal consorzio perché rilasciate con una licenza pubblica di libero utilizzo. In questo secondo caso, sarà necessaria una seria analisi delle licenze in entrata e una valutazione che possa portare a definire una licenza in uscita con esse compatibile. Il rapporto di "compatibilità" tra le principali licenze open data è esplicitato in una nota tabella contenuta nelle linee guida AgID, che è utile riportare.

Licenza opera derivata Licenza opera originaria	CC0	CC-BY	CC-BY-SA	IODL v. 2.0	IODL v. 1.0	ODbL
CC0						
CC-BY						
CC-BY-SA						
IODL v. 2.0						
IODL v. 1.0						
ODbL						

	La creazione di un'opera derivata e la sua pubblicazione è possibile
	La creazione di un'opera derivata potrebbe essere possibile ma vi è incertezza (ad esempio sui diritti licenziati) circa l'effettiva compatibilità o altri problemi (problema di stratificazione delle attribuzioni), oppure sul tipo di prodotto derivato (e.s. per la ODbL le modifiche dei dati sono rilasciabili solo con ODbL mentre i prodotti derivati come le mappe con ogni altra licenza).
	La creazione di un'opera derivata sotto la licenza proposta è impossibile

Come si può facilmente intuire, quanto diverse e disomogenee sono le licenze dei dati input tanto più saranno alte le probabilità di non riuscire a trovare una licenza unica per tutti gli output del progetto. Di conseguenza, in tutti quei casi in cui le licenze dei dati input non permettano il cambio di licenza sulle opere derivate, sarà necessario mantenere le stesse licenze anche sui dati output, non aggregandoli/miscelandoli in un unico dataset ma esponendoli come dataset ancora separati, ciascuno con i suoi metadati di licenza originari.

Ciò diventa ancor più stringente in tutti quei casi in cui le licenze in entrata negano la possibilità di fare "derivazioni" (esempio: licenze CC ND); questi dataset potranno essere ripubblicati solo così come sono stati acquisiti, senza essere disaggregati o miscelati con altri dataset.

Leggermente diverso è il discorso per le licenze che pongono particolari restrizioni sulle "derivazioni" (si veda il caso delle licenze tipo "Share Alike"); esse permettono di realizzare dataset derivati, ma pongono l'obbligo di applicare la stessa licenza anche sui dataset derivati. Qualora l'obiettivo fosse quello di realizzare ed esporre come output un dataset aggregato realizzato a seguito di una palese attività di derivazione, nascerebbero i classici problemi di compatibilità tra licenze. Per ovviare questi problemi diventa necessario mantenere separati gli outputs del progetto e redistribuirli come database non aggregati.

Al fine di ridurre al minimo la nascita (in fase output) di questioni di incompatibilità tra licenze, ove possibile, cioè nei casi in cui tra i partner del Progetto MISTRAL e gli enti che forniscono i dati input sussistano rapporti, si dovrebbe considerare l'ipotesi di intavolare un dialogo per portarli a modificare le licenze in modo da appianare a priori eventuali problemi di compatibilità.

(Avv. Simone Aliprandi)

