



Regione Lombardia

ALLEGATO A

AGENDA DIGITALE LOMBARDA

Processi di governance dell'Open Data

Novembre 2018



1 Scopo del documento

Questo documento delinea i processi necessari ad attuare il programma d'interventi previsto dall'area prioritaria "Valorizzare il patrimonio informativo pubblico" dell'Agenda Digitale Lombarda (DGR n.1887 del 30/05/2014) che riguarda sia Regione Lombardia sia gli Enti del SIREG.

In quest'area prioritaria Regione Lombardia si è data l'obiettivo di rendere disponibili le parti del proprio patrimonio informativo che lo consentono come Open Data, di favorire la pubblicazione sul portale www.dati.lombardia.it dei dati degli enti del territorio lombardo e di promuovere processi di aggiornamento automatico dei dati pubblicati.

La DGR XI/723 del 05/11/2018 "*Criteri per l'Open Data in Regione Lombardia*" ha aggiornato i criteri che impegnano la Giunta, il Sistema Regionale e gli Enti che aderiscono all'iniziativa ed utilizzano la piattaforma tecnologica www.dati.lombardia.it per la pubblicazione dei propri Open Data, precedentemente definiti dalla DGR IX/2904 dell'11-2-2012.

Questo documento sostituisce l'allegato "Piano di Governance Open data" approvato con il Decreto X/6115 del 10/07/2012.



Sommario

1	Scopo del documento	2
2	Contesto di riferimento	4
3	Il modello di Governance dell'Open Data in Regione Lombardia	5
3.1	Ruoli e attori.....	5
3.1.1	Team Open Data.....	5
3.1.2	Responsabile della banca dati (Data Owner)	6
3.1.3	Referenti Open Data nelle DG/DC e negli Enti del SIREG	6
3.1.4	Privacy Officer e Responsabile della Protezione dei Dati (RPD).....	6
3.2	Processo	7
3.2.1	Iniziativa di pubblicazione dei dataset	8
3.2.2	Verifica dei dataset da pubblicare	10
3.2.3	Pianificazione dei dataset da pubblicare.....	15
3.2.4	Pubblicazione dei dataset.....	15
3.2.5	Comunicazione, promozione dei dataset.....	17
3.2.6	Monitoraggio	17
3.3	Matrice di responsabilità.....	18
4	Acronimi.....	19
5	Riferimenti	20
5.1.1	Riferimenti Bibliografici Regione Lombardia	20
5.1.2	Sitografia.....	20



2 Contesto di riferimento

Le legislazioni comunitaria e nazionale valorizzano il patrimonio informativo pubblico quale strumento di trasparenza dell'attività amministrativa, diffusione della conoscenza nonché come elemento da impiegare per lo sviluppo economico e sociale del territorio. In questo contesto si inserisce il Programma Open Data di Regione Lombardia, nato nell'ambito dell'Agenda Digitale Lombarda nel 2012 che, a distanza di 6 anni, coinvolge sempre di più il territorio lombardo nella pubblicazione dei dati sul portale Open Data Lombardia (www.dati.lombardia.it) al fine di creare valore per cittadini, imprese e amministrazioni pubbliche.

Con il termine "Open Data" s'intende la pubblicazione sul Web dei dati delle pubbliche amministrazioni affinché siano liberamente accessibili a tutti, in formato aperto, senza restrizioni legali o altre forme di controllo che ne limitino la riproduzione; in Regione Lombardia, il settore pubblico raccoglie, produce, riproduce e diffonde un'ampia gamma di informazioni in molti settori di attività: ad esempio informazioni di tipo sanitario, sociale, economico, geografico, climatico, turistico, ambientale. Il portale Open Data Lombardia è lo strumento che permette oggi di rendere disponibili migliaia di dataset, suddivisi in 22 aree tematiche, oltre a funzionalità avanzate che consentono agli utenti di creare viste dei dataset per condividerle. Sul portale sono presenti dati rilevanti in ambito sanità, ambiente, energia. Da marzo 2012 a fine giugno 2018 si sono registrate circa due milioni di visualizzazioni e più di un milione di download.



3 Il modello di Governance dell'Open Data in Regione Lombardia

La gestione della pubblicazione d'informazioni sul Portale Open Data di Regione Lombardia richiede che siano identificati, all'interno dell'organizzazione regionale, precisi ruoli e attori (vedi par. 3.1), e che sia definito un processo con chiare responsabilità per ogni attività (vedi par. 3.2).

La Matrice di Responsabilità in coda al capitolo (vedi par. 3.3) fornisce una chiave di lettura sintetica dell'intero processo indicando, per ciascun ruolo identificato, i differenti livelli di responsabilità sulle azioni previste.

3.1 Ruoli e attori

Sono identificati i seguenti ruoli:

- 1) Team Open Data;
- 2) Responsabile della banca dati (Data Owner);
- 3) Referenti Open Data nelle DG/DC e negli Enti del SIREG;
- 4) Privacy Officer e RPD

Ciascun ruolo è chiamato a lavorare in stretto collegamento e con coordinazione puntuale con gli altri; è possibile che uno stesso soggetto sia incaricato di più ruoli o che gli sia chiesto di fare da punto di contatto tra differenti ruoli.

Di seguito sono descritti i ruoli e le strutture che ne fanno parte.

3.1.1 Team Open Data

Il Team Open Data svolge sia un ruolo di regia del programma Open Data che un ruolo di governo tecnico. In particolare:

- definisce gli obiettivi e ne verifica il raggiungimento;
- approva il piano degli interventi ed esegue il monitoraggio delle attività;
- verifica la coerenza delle iniziative con il programma regionale;
- approva le regole per la comunicazione e la promozione;
- ha il controllo diretto delle attività di sviluppo del progetto;
- coordina il rapporto tra strutture tecniche, DG/DC ed Enti del SIREG;
- definisce i processi e gli standard tecnici da applicare.

Fanno parte della Team Open Data:

- il Dirigente della struttura regionale che ha responsabilità sul programma Open Data (al momento il Dirigente della Struttura Semplificazione e Trasformazione Digitale) e suoi collaboratori;
- il responsabile, in Lombardia Informatica, del programma Open Data (al momento il Responsabile della Funzione Innovazione Digitale) e suoi collaboratori.



3.1.2 Responsabile della banca dati (Data Owner)

I Data Owner sono i responsabili del procedimento amministrativo che popola la specifica fonte del dato, che ne cura la qualità e il relativo aggiornamento; essi sono tipicamente Dirigenti di Regione Lombardia o degli Enti SIREG che coordinano un gruppo di persone che svolgono il loro lavoro quotidiano attorno alla fonte del dato e hanno **la responsabilità di decidere** quali dataset pubblicare, verificare vincoli (es. normativi) e qualità dei dati, pubblicarli e tenerli aggiornati con il supporto del personale incaricato nella propria struttura.

3.1.3 Referenti Open Data nelle DG/DC e negli Enti del SIREG

All'interno di ciascuna DG/DC e di ciascun Ente SIREG dovranno essere coinvolti – ognuno con il proprio ruolo – diversi attori, e principalmente:

- il Direttore Generale, che eventualmente delega i responsabili dei diversi procedimenti all'identificazione e all'analisi dei dati esportabili in formato open per la sua Direzione (vedi par. 3.1.2);
- i referenti informatici che supportano i responsabili dei procedimenti.

Per semplificare i processi e la comunicazione, deve essere formalmente individuato, in ciascuna Direzione o Ente, un **Referente Open Data** che farà da figura unica di riferimento. Tale figura potrà essere scelta tra i responsabili dei procedimenti o tra i loro referenti informatici.

Il Team Open Data interagirà principalmente con il Referente Open Data per tutte le attività d'identificazione, analisi e pubblicazione dei dataset, e lo supporterà nelle seguenti attività specifiche:

- aggiornare le competenze in merito alle norme ed alle Linee Guida nazionali emanate da AgID e/o dal Team Digitale in ambito Open Data;
- individuare – con i responsabili dei diversi procedimenti – quali dati potrebbero essere pubblicati in modalità Open Data;
- proporre i dataset individuati al Team Open Data;
- ottenere – ove necessario – le autorizzazioni formali alla pubblicazione dei dataset individuati e analizzati (vedi par. 3.2.2);
- predisporre – con i referenti informatici ed i responsabili dei procedimenti - i metadati e le schede descrittive associati a ciascun dataset;
- definire, in stretto raccordo con il team Open Data, le modalità di pubblicazione;
- supervisionare il processo di aggiornamento dei dataset.

3.1.4 Privacy Officer e Responsabile della Protezione dei Dati (RPD)

Il Privacy Officer, che presidia le procedure e le attività in attuazione del Regolamento (UE) 2016/679 in materia di protezione dei dati personali, ha il compito di fornire indicazioni al titolare sulla pubblicazione dei dataset contenente dati personali di cui Regione Lombardia è Titolare del trattamento.



Il Responsabile della Protezione dei Dati (RPD) ha il compito di informare e fornire consulenza al titolare del trattamento o al responsabile del trattamento, nonché ai dipendenti che eseguono il trattamento, in merito agli obblighi derivanti dal GDPR.

3.2 Processo

Aprire i dati a disposizione dell'Amministrazione comporta una serie di passaggi che è opportuno strutturare in maniera esplicita attraverso un processo organizzato che prenda in considerazione le diverse variabili esistenti.

L'approccio per processi qui descritto, declinato nell'attività di apertura dei dati dell'Amministrazione, sottolinea l'importanza di:

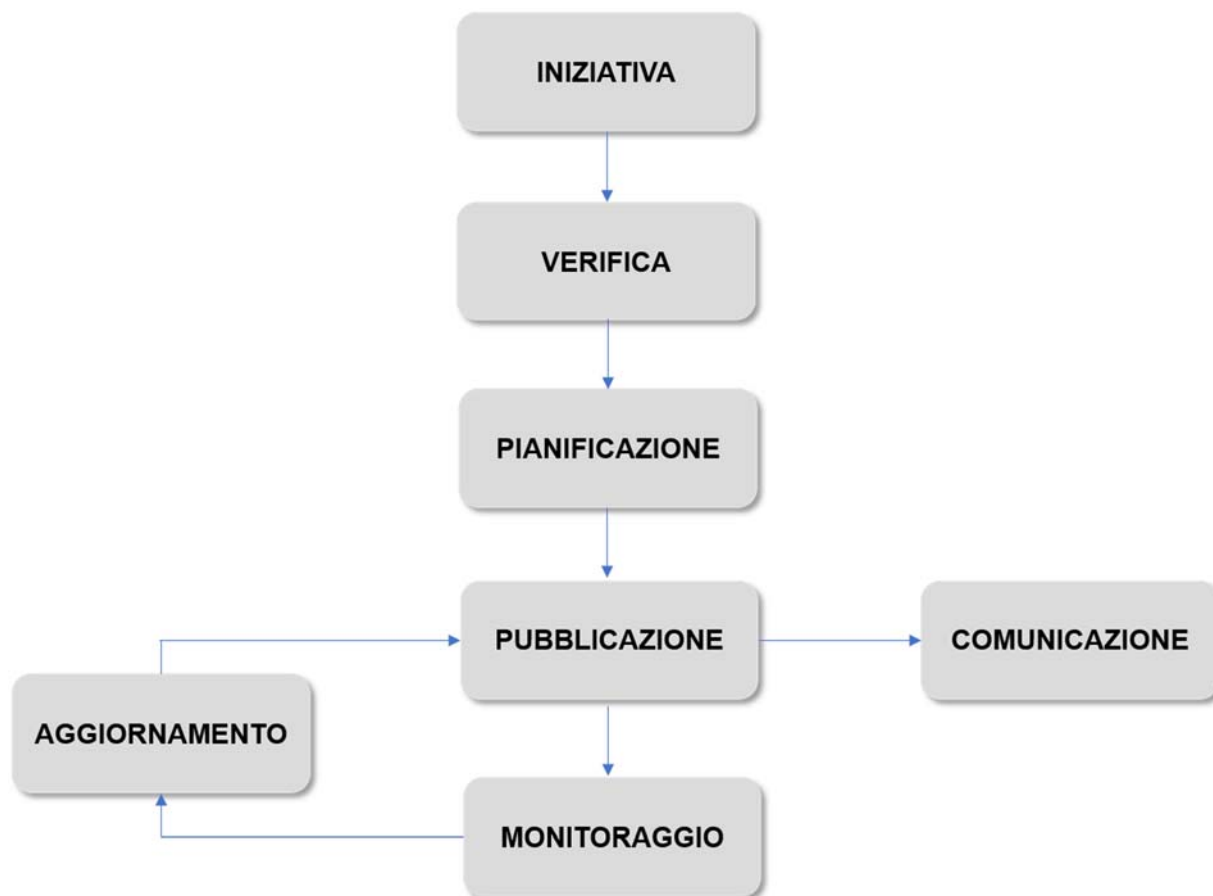
- aver chiaro i requisiti di partenza, nelle diverse fasi connesse all'apertura dei dati;
- valutare le attività previste dai processi;
- conseguire risultati misurabili in termini di efficacia complessiva dell'azione di apertura dei dati;
- ottenere un miglioramento continuo dei processi, basato su misurazioni oggettive definite dal Team Open Data.

Descrivere i processi connessi all'apertura dei dati a disposizione di un'Amministrazione Pubblica vuol dire, quindi, identificare e descrivere la sequenza strutturata di azioni che sottendono tale attività, partendo dall'identificazione del dato da pubblicare e passando per la sua elaborazione, per arrivare alla pubblicazione e diffusione attraverso i canali più idonei.

Spostandosi da una visione complessiva e sistemica a una maggiormente operativa e analitica, a valle della decisione di aprire un insieme di dati in possesso dell'Amministrazione per metterli a disposizione del cittadino e della comunità, è possibile identificare un processo strutturato nelle seguenti fasi:

1. Iniziativa di pubblicazione dei dataset (ricerca basi dati da cui estrarre dataset potenzialmente pubblicabili)
2. Verifica dei dataset da pubblicare
3. Pianificazione dei dataset da pubblicare
4. Pubblicazione dei dataset
5. Comunicazione e promozione dei dataset pubblicati
6. Monitoraggio dell'efficacia
7. Aggiornamento dei dataset pubblicati

Descrizione del “Ciclo di vita” del dataset e delle attività di gestione del dato:



Di seguito sono dettagliate le singole fasi del processo.

3.2.1 Iniziativa di pubblicazione dei dataset

Questa fase prevede l'identificazione dei dati da pubblicare in base a un'analisi che risponda alle seguenti domande:

1. quali sono i dati a disposizione dell'Amministrazione?
2. quali sono i dati d'interesse per la Comunità?

3.2.1.1 Quali sono i dati a disposizione dell'Amministrazione?

L'Amministrazione Pubblica, come qualsiasi organizzazione complessa, gestisce quotidianamente una grande quantità di dati che servono per garantire l'operatività e l'erogazione dei servizi al cittadino. Con lo sviluppo dell'Information & Communication Technology (ICT) e la diffusione dei sistemi di e-Government, tali dati sono ormai disponibili in formato digitale e possono quindi essere condivisi e diffusi. Per farlo, però, è necessario identificarli nell'ambito della struttura amministrativa dell'Ente.



Definire dove sono i dati a disposizione, vuol dire:

- identificare quali sono le strutture funzionali dell'Amministrazione che detengono dati potenzialmente utili;
- comprendere, in ciascun ambito, quali sono le fonti di dati di rilievo; in altri termini, è necessario **identificare archivi strutturati, elenchi, basi dati** prodotte da software in uso presso l'amministrazione.

Le organizzazioni complesse non sempre sono a completa conoscenza dell'intero corpus di dati che producono o che gestiscono. In molti casi, infatti, il dato è il risultato di un processo o il sottoprodotto di un'elaborazione, funzionale ad altri scopi.

In generale, è possibile identificare queste tipologie di dati in possesso dell'Amministrazione:

- **dati necessari alla gestione delle attività:** sono tutti quei dati che servono all'Amministrazione per il corretto funzionamento dei suoi processi. Sono i dati in ingresso (o come abbiamo già citato, in input) che sorreggono e riforniscono i processi gestiti dall'Ente;
- **dati prodotti come risultato dell'attività:** sono tutti quei dati che l'Amministrazione produce, spesso alla conclusione di un iter procedurale "monolitico", nell'esercizio delle sue funzioni. Sono i dati in uscita (output) ottenuti come risultato dei processi e dei procedimenti gestiti dall'Ente;
- **dati prodotti nella gestione delle attività:** sono quei dati che l'Amministrazione ottiene o produce come sottoprodotto di un processo, che derivano da lavorazioni intermedie o che sono prodotti con la finalità di essere dati in pasto ad altri procedimenti.

Per ottimizzare e rendere efficiente il sistema, è necessario che ogni responsabile di procedimento all'interno della DG/DC o Ente SIREG analizzi i dati che gestisce/produce, con il supporto di chi se ne occupa operativamente, al fine di valutarne la possibile apertura eventualmente dopo averli sottoposti a elaborazioni, quali anonimizzazione o aggregazione, in caso alcune informazioni siano protette da privacy (vedi par.3.2.2).

3.2.1.2 *Quali sono i dati d'interesse per la Comunità?*

Nella La DGR XI/723 del 05/11/2018 che ha ri-definito i "Criteri per l'Open Data in Regione Lombardia" è suggerito di *"adottare criteri che tengano conto della potenziale utilità dei dati nella creazione di valore sociale e di mettere in atto modalità che possano permettere ai potenziali riutilizzatori di esprimere il proprio interesse, al fine di attuare una politica di apertura sempre più guidata dalla domanda. Allo scopo, sul portale Open Data, è presente in prima pagina la funzione "Suggerisci un dataset", che permette a chiunque di richiedere la pubblicazione di un dato"*.

I possessori dei dati sono invitati ad individuare iniziative che possano essere utili a coinvolgere sia gli stakeholder interni all'amministrazione che quelli esterni (es. (e.g., studenti universitari, soggetti preposti a indagini e analisi statistiche e/o economiche, datajournalist, startup e aziende).



È tuttavia necessario riconoscere che non essendo semplice valutare l'interesse che susciterà un determinato dataset, si suggerisce, quando un dato è pubblicabile senza costi significativi, di provvedere alla sua 'apertura' anche nel momento in cui non se ne ravveda un'utilità immediata.

3.2.2 Verifica dei dataset da pubblicare

Una volta identificate le possibili fonti di dataset, è necessario procedere ad analisi più approfondite per valutare l'effettiva "pubblicabilità" dei dati come Open Data.

Questa attività deve essere svolta da chi sa come e dove i dati sono gestiti fisicamente (referente informatico della DG/DC o dell'Ente SIREG), in cooperazione con chi conosce il dato e le sue funzionalità, il responsabile del procedimento.

Per ogni singolo dataset potenziale occorre:

1. verificare se i dati sono distribuibili sotto il profilo giuridico;
2. analizzare la qualità dei dati;
3. definire i processi di produzione del dataset;
4. produrre documentazione di supporto.

3.2.2.1 Verificare se i dati sono distribuibili sotto il profilo giuridico

In caso di pubblicazione di dati personali Regione Lombardia può essere Titolare del trattamento o Responsabile del trattamento.

In qualità di **Titolare del trattamento** dei dati determina le finalità e i mezzi del trattamento (art. 4, punto 7 GDPR) nel rispetto del combinato disposto tra art. 5, 6 e considerando n. 50 del GDPR.

Ciò significa che i **Data Owner**, con il supporto del Privacy Officer e del Responsabile della Protezione dei Dati (RPD), **decidono** se è opportuno pubblicare dati personali previa **valutazione sulla liceità e necessità** del trattamento, secondo le indicazioni della La DGR XI/723 del 05/11/2018 "*Criteria per l'Open Data in Regione Lombardia*" (articolo specifico 4.6 Privacy). A questa valutazione sono legate anche la scelta della **licenza** di utilizzo dei dati e della «**persistenza**» dei dati (ovvero l'intervallo di tempo in cui è lecito che ciascun dataset resti pubblicato), che devono essere indicate nella scheda metadati di ciascun dataset.

Qualora, invece, Regione Lombardia ospiti sul proprio Portale Open Data dataset contenenti dati personali di cui un altro **Ente terzo** è Titolare del trattamento, essa deve essere nominata dall'Ente terzo – con specifico atto di nomina – **Responsabile del trattamento** per poter trattare i dati per conto del Titolare.

In tal caso, **la valutazione sulla liceità e necessità** del trattamento dei dati personali **resta in capo all'Ente terzo Titolare del trattamento**, che dovrà comunicare di volta in volta al Team Open Data l'esito della valutazione, e assumersi la responsabilità della pubblicazione e delle scelte in merito alla licenza d'uso ed alla persistenza dei dati.



3.2.2.2 *Analizzare la qualità dei dati*

Dopo diversi anni di esperienze in ambito Open Data, anche in Italia, si è compreso che la qualità dei dati è un aspetto fondamentale da curare e che può incidere significativamente sulla probabilità di riuso da parte degli stakeholder.

È quindi importante eseguire verifiche qualitative sui dati, prima della loro pubblicazione e mettere in atto eventuali azioni correttive laddove necessario.

Gli aspetti fondamentali di cui tenere conto sono:

- **Accuratezza sintattica** - Il formato dei dati dovrebbe essere controllato perché sia coerente con la realtà (es. una data dovrebbe essere riportata sempre nel formato corretto);
- **Accuratezza semantica** - Il valore dei dati deve essere il più possibile affidabile: ciò si ottiene utilizzando soprattutto delle fonti interne o esterne certificate e/o controllate. La misura di tale parametro di qualità si basa sulla considerazione che l'accuratezza semantica è tanto maggiore quanto più:
 - il processo che genera il dato prevede un controllo;
 - la fonte è affidabile;
 - la fonte certifica il dato secondo le normative specifiche (es. il Ministero delle Finanze certifica il codice Fiscale, l'ISTAT i dati demografici).
- **Tempestività** – È importante garantire che l'aggiornamento del **dataset pubblicato in Open Data avvenga in tempi congrui** con quelli di aggiornamento dell'informazione reale. **Il livello adeguato di aggiornamento** dipende dal tipo di dato: possono esserci dati che hanno bisogno di essere aggiornati molto frequentemente e altri per i quali è sufficiente una frequenza di aggiornamento periodica. Viene calcolata come il rapporto tra il periodo che passa tra due aggiornamenti del dataset e la periodicità stimata di variazione delle informazioni reali.
- **Completezza** – Occorre valutare **la completezza del dato e comunicarla**: viene calcolata come la percentuale di rappresentatività del dataset rispetto all'universo complessivo reale (ad esempio: il dataset biblioteche pubbliche della Lombardia' quante biblioteche contiene rispetto al totale?)

La pretesa di ottenere la perfezione assoluta non è realistica, anche perché richiederebbe dei costi eccessivi rispetto ai benefici, ma occorre verificare **se sono necessarie azioni correttive** per migliorare la bontà del dato, con la consapevolezza che questo intervento può essere realizzato in modalità graduale e progressiva.

È importante comunque giudicare se si è raggiunto un **livello sufficiente di correttezza, aggiornamento e completezza del dato** da consentire a enti terzi elaborazioni utili; tale livello dipende dal tipo di dato e dall'universo trattato.

È inoltre ampiamente dimostrato dall'esperienza come la scelta di aprire il dataset possa essere funzionale a migliorarne la qualità, anche attraverso processi di coinvolgimento della Comunità di riferimento. Sono numerosi i casi di segnalazioni di incompletezza o non accuratezza del dato che hanno favorito un aumento della qualità dei dati.



3.2.2.3 *Definire i processi di produzione del dataset*

Dopo aver eseguito l'analisi della qualità del dataset, si passa all'analisi da parte dei referenti tecnici che valutano prima di tutto la complessità ed il tempo necessario per gli interventi utili ad estrarre i dati ed eventualmente creare un "job" di pubblicazione/aggiornamento automatico.

Quest'attività tecnica necessita della stretta collaborazione tra il Data Owner (che conosce i processi di produzione del dato) e le componenti tecniche del Team Open Data, che conoscono le funzionalità del portale Open Data e sono in grado di mettere a disposizione strumenti e buone pratiche.

In generale, la stima dei costi e dei tempi per la produzione del dataset, è influenzata da una serie di fattori:

- 1) **La fonte dei dati è identificata?** Si deve sapere dove sono fisicamente collocate le basi dati su cui si va a lavorare, chi ne ha la gestione, quali tecnologie sono in uso.
- 2) **La struttura dei dati è nota?** I moderni data base hanno una struttura che può essere anche molto complessa; conoscerla è ruolo di chi la gestisce ed è necessario per estrapolare i dati che si intende pubblicare.
- 3) **Sono necessarie trasformazioni dei dati?** Si è già accennato al fatto che talvolta si può decidere che il dataset sia composto di campi che sono a loro volta risultato di alcune elaborazioni, la cui complessità può essere variabile e incide sui tempi di produzione del dataset. Per fare alcuni esempi di possibili interventi che possono rendersi necessari, ecco alcune domande che chi vuole esportare il dataset dovrà porsi:
 - a. **esistono già strumenti con funzionalità di esportazione dei dati?** In molti casi sono già state previste tali funzioni, in altre occorreranno interventi per realizzarle;
 - b. **il dato è geo-localizzato o deve esserlo?** Se una base dati contiene già le coordinate per geo-localizzare l'informazione è cosa utile poterla esportare. In alcuni casi si hanno solo gli indirizzi che possono però anche essere utilizzati per geo-localizzare il dato; questa però è un'attività che va preventivata e può richiedere l'impegno anche del gestore del dato che deve intervenire per i casi in cui l'indirizzo incompleto non permette una corretta geo-localizzazione.;
 - c. **ci sono dati che devono essere resi anonimi?** Se sì, è necessario, prima che il dato sia trasmesso per la sua pubblicazione, prevedere l'intervento con opportune funzioni che trasformano in dati anonimi le informazioni;
 - d. **alcuni dati devono essere di sintesi?** Se non esiste già una funzione che restituisce dati aggregati o di sintesi dai dati originari, questa dovrà essere realizzata.
- 4) **Ogni quanto tempo il dataset deve essere aggiornato?** Questo dipende dalla frequenza di aggiornamento della base dati di origine. Considerando che i dataset sono immagini statiche del patrimonio informativo devono essere previsti congrui tempi di aggiornamento; per basi dati che hanno aggiornamenti periodici (es.: rilevazioni annuali), potrà essere previsto l'aggiornamento del dataset di conseguenza; nel caso invece di basi dati che subiscono aggiornamenti in modo costante, si dovrà prevedere una congrua tempistica di aggiornamento del dataset. Tutte le tempistiche vanno opportunamente indicate nei metadati (vedi par. successivo). È importante considerare che una volta messo in linea il



dataset questo dovrà essere tenuto aggiornato e va deciso anche se aggiornare sostituendo il vecchio o mantenendo la serie storica dei dataset.

3.2.2.4 *Produrre documentazione di supporto*

Nel corso dell'analisi dovrà essere predisposta un'opportuna documentazione che sarà allegata ai dataset al momento della pubblicazione.

Scheda Metadati

La scheda metadati è il "core" della documentazione del dataset. Essa è necessaria, poiché permette ai fruitori dei dati di comprendere la loro natura e di come utilizzarli, sia dal punto di vista informatico che del contenuto informativo.

La compilazione della scheda deve essere attuata in sinergia tra il personale che conosce profondamente il dato e le sue funzionalità (attraverso il Referente Open Data) e il supporto tecnico (Service Manager e/o referente informatico di DG/DC o dell'Ente SIREG) che sa come e dove i dati sono gestiti fisicamente.

Di seguito è presentata la scheda metadati individuata con una breve descrizione delle informazioni richieste; la compilazione deve avvenire con il massimo dettaglio possibile e in particolare per i campi colorati in verde.

Sezione	Campo	Valore
Informazioni interne	Stato validazione scheda	
Dati identificativi	Titolo	Nome del dataset
	Ente fornitore	Titolare del dato
	Data pubblicazione	Data in cui il dataset è reso disponibile on line
	Data di ultima modifica dato	Data di ultima modifica del dataset
	Frequenza di aggiornamento	Periodicità con cui il dataset è aggiornato (in giorni o mesi)
	Persistenza	Data di cancellazione del dataset (obbligatorio per tutti i dataset, non solo in caso di presenza di dati personali)
Informazioni generali	Formato di pubblicazione	Quale formato (standard) è utilizzabile per la pubblicazione
	Tipologia dati	Di dettaglio o statistici
	Descrizione breve	Questa compare nella schermata dei metadati on-line



	Descrizione estesa	Questa è più approfondita e può comparire in un documento allegato
	Parole chiave	Utilizzabili da motori di ricerca e per la classificazione
	Copertura geografica	A quale contesto territoriale si riferiscono i dati
	Modalità di raccolta	Es: rilevazione on-line o manuale
	Link documentazione tecnica	Indirizzo web a pagine di descrizione tecnica
	URL sito (eventuale)	Indirizzo web dove i dati sono presentati tramite altri servizi (es.: portale istituzionale)
Riferimenti per informazioni	DG/DC/Ente SIREG	Riferimento organizzativo di Regione Lombardia
	Unità organizzativa	Riferimento organizzativo di Regione Lombardia
	Dirigente Responsabile	Riferimento organizzativo di Regione Lombardia
	Struttura	Riferimento organizzativo di Regione Lombardia
	Dirigente Responsabile	Riferimento organizzativo di Regione Lombardia
Informazioni su protezione dei dati personali	Trattamento dei dati	<ul style="list-style-type: none"> • Non sono presenti dati personali oppure • Verificata la necessità e liceità del trattamento
Informazioni su distribuzione, limitazioni d'uso e accessibilità	Licenza	CC-BY 4.0 (o altra solo in caso di presenza di dati personali)
Informazioni sulla qualità	Modalità di realizzazione del prodotto	Fonte (processo, soggetti coinvolti)
	Percentuale di copertura	Completezza
	Tempestività	Rapporto tra la periodicità media di variazione dell'informazione contenuta nel dataset e periodo che passa tra due aggiornamenti dei dati pubblicati



Scheda descrizione dataset

Se i metadati comunicano all'utente informazioni di base sul dataset, sulla base delle quali esso decide se questo può essere di sua utilità, per favorire l'utilizzo del dataset può essere certamente utile descriverne i contenuti.

La scheda di descrizione del dataset dà all'utente dettagli su com'è organizzato; possibili contenuti della scheda sono:

- descrizione dettagliata del dataset, da dove è originato, per quali scopi è realizzato;
- legenda dei nomi delle colonne (es. IDEXT = Identificativo Esterno);
- tipo del campo (es.: importo=intero lungo due decimali);
- legenda di sigle, acronimi o altre abbreviazioni utilizzate nei contenuti (es. 1=Si, 0=No).

Queste informazioni – in particolare la legenda sul significato dei nomi delle colonne – sono fondamentali per un corretto utilizzo dei dataset in seno ad applicazioni sviluppate da soggetti terzi, e devono essere prodotte da chi gestisce la base dati in collaborazione con la parte tecnica.

3.2.3 Pianificazione dei dataset da pubblicare

I dataset individuati e verificati possono essere predisposti e pubblicati in autonomia dai Data Owner oppure con il supporto del Team Open Data.

In caso di pubblicazione autonoma da parte del Data Owner, non occorre una vera e propria pianificazione ma è necessario comunque informare il Team Open Data.

Nel caso in cui il Data Owner si affidi al Team Open Data per la pubblicazione, è necessaria una pianificazione. In questo caso, sulla base delle informazioni raccolte in fase di verifica, il Team Open Data farà una stima dei giorni uomo previsti per le attività di pubblicazione del dataset (incluso la creazione di eventuali "job" automatici) ed una previsione di quando potrà essere realizzata in base alle attività in corso in quel periodo.

Eventuali urgenze devono essere concordate con il team Open Data.

NOTA BENE

Nel caso il dataset contenga dati personali, il titolare del trattamento (Data Owner interno a regione Lombardia o Ente terzo) deve, prima di pubblicare il dato o di chiederne la pubblicazione al team Open Data, inoltrare al Team Open Data una comunicazione nella quale dichiara di aver valutato i presupposti di liceità e necessità e deve assumersi la responsabilità della pubblicazione e delle scelte in merito alla licenza d'uso ed alla persistenza dei dati.

3.2.4 Pubblicazione dei dataset

Solo dopo le fasi di iniziativa, verifica e pianificazione è dato il via all'effettiva estrazione e pubblicazione del dataset sul portale Open Data anche se alcune delle fasi di estrazione possono essere anticipate, per dare supporto alle fasi di iniziativa e verifica.



Queste attività sono svolte dalla parte tecnica che deve, comunque, restare in contatto con chi gestisce le basi dati coinvolte nell'intervento e con il Team Open Data, sia per avere un supporto decisionale, sia per dare riscontro sull'avanzamento dell'attività.

Le principali attività per la pubblicazione sono riassunte nei paragrafi seguenti.

3.2.4.1 *Definizione del dataset*

Partendo da quanto stabilito in analisi, si deve a questo punto stabilire con precisione quali informazioni saranno esportate, individuando quali dati e in che formato, comunque aperto, è possibile esporli (ad es.: formato tabellare, shape file, etc).

3.2.4.2 *Integrazione della documentazione*

La tabella dei metadati e la scheda di descrizione, precedentemente predisposti (vedi Paragrafo 3.2.2.4), vanno in questa fase verificati ed eventualmente integrati.

3.2.4.3 *Estrazione*

Qualora i dati siano gestiti da una base dati informatizzata, saranno realizzati gli strumenti che permettono di metterli a disposizione del Team Open Data (es. viste sul DB, ETL, etc) concordando modalità e tempi con il Team Open Data. Qualora i dati siano gestiti con strumenti personali (es. tabelle elettroniche o DB residenti su PC), dovranno essere concordate con il Team Open Data le modalità di fornitura dei dati.

3.2.4.4 *Geolocalizzazione*

La geolocalizzazione delle informazioni ne facilita il riutilizzo all'interno delle applicazioni che fanno uso di mappe; se però non è realizzata alla fonte è richiesto un lavoro d'integrazione che può essere fatto anche in un momento successivo alla prima pubblicazione del dataset, ma del quale occorre tenere conto in termini di tempo e dell'impegno organizzativo.

È il Referente Open Data che, sentiti i responsabili dei dati all'interno delle DG/DC o nell'Ente SIREG, richiede questa attività al Team Open Data.

Perché sia possibile procedere con la geolocalizzazione è necessario che nei dati siano presenti "indirizzo e numero civico" il più possibile "normalizzati" ovvero inseriti in campi separati. Occorre che il Data Owner consideri che, siccome è probabile che il processo di geolocalizzazione produca una serie di errori, sarà necessario un suo intervento per la gestione dei casi non gestibili in automatico.

3.2.4.5 *Pubblicazione sul portale Open Data*

Ovvero invio all'ambiente ospitante, del dataset e di tutte le informazioni predisposte corredo.

È da rilevare che molte di queste azioni, sono da realizzare sia alla prima pubblicazione del dataset bersaglio sia da reiterare nel tempo per i periodici aggiornamenti, anche perché è possibile che il dataset sia sottoposto a modifiche e integrazioni.



3.2.5 Comunicazione, promozione dei dataset

È importante che la pubblicazione dei dataset sia accompagnata da un'attività di comunicazione e promozione, che sono seguite dal Team Open Data, in collaborazione con la Funzione Comunicazione di Regione Lombardia.

Anche i responsabili dei dati sono chiamati a informare i propri interlocutori più diretti delle attività che stanno compiendo in chiave Open Data, e a coordinarsi con il Team Open Data per interventi di carattere straordinario.

3.2.6 Monitoraggio

Il Team Open Data garantisce un costante monitoraggio al quale tuttavia sono chiamati a concorrere tutte le strutture coinvolte a vario titolo nel programma Open Data.

Attraverso strumenti automatici il team Open Data monitorerà costantemente la qualità dei dati e segnalerà ai titolari dei dati eventuali anomalie quali ad esempio la non completezza dei metadata, il non rispetto dei vocabolari controllati o il mancato aggiornamento dei dati nei tempi stabiliti.

Il Team Open Data, sempre con strumenti automatici, garantirà la cancellazione dei dataset al termine del periodo di persistenza indicato dal data Owner.

Inoltre, in considerazione del fatto che i dati pubblicati creano valore solo se qualcuno li usa, sarà analizzato l'interesse che suscitano i diversi dataset e, se possibile, tracciate le applicazioni che ne fanno uso.

Il Team Open Data estrarrà dal portale Open Data informazioni quali il conteggio dei download e del numero di visualizzazioni del dataset, in grado di fornire un primo indicatore dell'interesse per un determinato set di dati.

L'identificazione dei casi di riuso, considerato che l'accesso ai dati è per lo più anonimo e non tracciabile, è particolarmente difficile. Tuttavia, dall'analisi dei "referral" e dalle interazioni con gli utenti che scrivono all'amministratore del portale, è possibile intercettare alcuni casi d'uso che saranno censiti in un apposito dataset al fine di darne evidenza pubblica.

3.3 Matrice di responsabilità

Come chiave di lettura del processo di governance ed elemento di correlazione tra le diverse attività e i ruoli sopra definiti, si presenta la Matrice di Responsabilità. Tale tabella mostra in modo sintetico le aree d'interazione e i differenti livelli di responsabilità tra diversi ruoli nelle attività previste per gli Open Data. Il modello utilizzato è il RACI, che prevede che per ogni associazione ruolo/fase del processo si definisca un livello di responsabilità del ruolo, individuato tra i seguenti:

- R: Responsabile
- A: Approva
- C: Consultato
- I: Informato

	Team Open Data	Data Owner	Referenti Open Data nelle DG/DC e nel SIREG	Privacy Officer o RPD
1. Iniziativa	A	R	C	I
2. Verifica	I	R/A	C	C ¹
3. Pianificazione	R/A	I	C	I
4. Pubblicazione	A	R	C	I
5. Comunicazione, promozione	R/A	I	I	I
6. Monitoraggio	R/A	I	I	I
7. Aggiornamento	A	R	C	I

¹ Il Privacy Officer e/o il RPD vengono consultati unicamente se sono presenti dati personali



4 Acronimi

AgID	Agenzia per l'Italia Digitale
Dataset	Insieme dei dati strutturato e pubblicato in formato aperto
DB	Data Base
DG/DC	Direzioni Generali/Direzioni Centrali di Regione Lombardia
DGR	Delibera Giunta Regionale
ETL	Extract, Transform, Load
ICT	Information and Communication Technology
Licenza CC-BY 4.0	Creative Commons Attribution 4.0 International Public License
SIREG	Sistema Regionale



5 Riferimenti

5.1.1 Riferimenti Bibliografici Regione Lombardia

- [1] Regione Lombardia, D.G.R. IX/2595 del 30 novembre 2011 (Agenda Digitale Lombarda)
- [2] Regione Lombardia, D.G.R. IX/2904 del 11 gennaio 2012 (Criteri generali per l'Open Data)
- [3] Regione Lombardia, L.r. 7 del 18 aprile 2012 (Misure per la crescita, lo sviluppo e l'occupazione)

5.1.2 Sitografia

- [1] Sito Web di Regione Lombardia www.regione.lombardia.it
- [2] Portale Open Data di Regione Lombardia www.dati.lombardia.it
- [3] Sito web di AgID www.agid.gov.it