REGIONE LAZIO

REGIONE LAZIO

Direzione: LAVORI PUBBLICI E INFRASTRUTTURE, INNOVAZIONE TECNOLOGICA

Area: SISTEMI INFORMATIVI TRASVERSALI, INTEROPERABILITA', OPEN DATA E PRIVACY

DETERMINAZIONE *(con firma digitale)*

N. G11119 **del** 21/08/2024 **Proposta n.** 30794 **del** 16/08/2024

Oggetto:

Approvazione della versione definitiva delle Linee Guida Regionali Open Data.

Proponente:

Estensore	TEODORI VALERIO	_____*firma elettronica*_____
Responsabile del procedimento	D'AMBROSIO MARIA ROSARIA	_____*firma elettronica*_____
Responsabile dell' Area	A. CORBELLI	_____*firma digitale*_____
Direttore Regionale	L. MARTA	_____*firma digitale*_____

Firma di Concerto

Copia

Richiesta di pubblicazione sul BUR: SI

OGGETTO: Approvazione della versione definitiva delle Linee Guida Regionali Open Data.

IL DIRETTORE DELLA DIREZIONE REGIONALE LAVORI PUBBLICI E INFRASTRUTTURE, INNOVAZIONE TECNOLOGICA

SU PROPOSTA del Dirigente dell'Area "Sistemi informativi trasversali, interoperabilità, open data e privacy";

VISTA la Legge Costituzionale 18 ottobre 2001, n. 3;

VISTO lo Statuto della Regione Lazio, approvato con legge statutaria 11 novembre 2004, n. 1;

VISTO il Decreto Legislativo 30 marzo 2001, n. 165 "Norme generali sull'ordinamento del lavoro alle dipendenze delle amministrazioni pubbliche" e s.m.i.;

VISTA la legge regionale 18 febbraio 2002, n. 6 "Disciplina del sistema organizzativo della Giunta e del Consiglio regionale e disposizioni relative alla dirigenza ed al personale regionale", in particolare l'art. 11, commi 1 e 1-ter che prevedono, tra l'altro, che per l'esercizio delle attività amministrative e gestionali, il sistema organizzativo della Giunta è costituito dalla Direzione generale, a cui è preposto il Direttore generale;

VISTO il regolamento regionale 6 settembre 2002, n. 1 "Regolamento di organizzazione degli uffici e dei servizi della Giunta Regionale" e successive modifiche ed integrazioni, in particolare l'articolo 22 "Criteri per la definizione dell'assetto organizzativo e per l'istituzione delle aree, degli uffici e delle posizioni dirigenziali individuali";

VISTO il Regolamento Regionale 23 ottobre 2023 n. 9 e s. m. i, che istituisce la Direzione regionale "Lavori pubblici e infrastrutture, innovazione tecnologica", ed in particolare l'art. 9, co. 2, ai sensi del quale la Direzione regionale è resa operativa dalla data di sottoscrizione del contatto del Direttore regionale incaricato;

VISTA la deliberazione di Giunta regionale n. 9 dell'11 gennaio 2024, con la quale è stato conferito l'incarico di Direttore della Direzione regionale "Lavori pubblici e infrastrutture, innovazione tecnologica" all'Ing. Luca Marta;

PRESO ATTO che il contratto dell'incarico di cui al punto precedente è stato sottoscritto in data 1° febbraio 2024;

VISTE le Direttive del Direttore generale prot. 1414222 del 05 dicembre 2023 e prot. 132306 del 30.01.2024 emanate in attuazione della riorganizzazione dell'apparato amministrativo di cui al regolamento regionale 23 ottobre 2023, n. 9 e s. m. i;

Copia

VISTA la V Direttiva del Direttore generale prot. n. 231852 del 19 febbraio 2024 che rettifica parzialmente quanto disposto con le precedenti direttive, relativamente alle competenze transitate;

VISTO l'Atto di Organizzazione n. G01353 del 12.02.2024 avente ad oggetto "Organizzazione della Direzione regionale "Lavori pubblici e infrastrutture, innovazione tecnologica" con il quale, in attuazione di quanto stabilito dalla Direttiva prot. n. 132306 del 30 gennaio 2024, si è provveduto a definire l'assetto organizzativo della Direzione, confermando e istituendo le strutture organizzative a rilevanza dirigenziale, secondo le competenze transitate in tutto o in parte a seguito del processo di riorganizzazione e i successivi Atti di organizzazione di modifica n. G01861 del 22 febbraio 2024, n. G02255 del 29 febbraio 2024, G04321 del 16 aprile 2024;

VISTO l'Atto di Organizzazione n. G17285 del 7 dicembre 2022 del Direttore regionale della Direzione regionale "Affari Istituzionali e Personale" con il quale è stato conferito l'incarico di Dirigente dell'Area "Sistemi informativi trasversali, interoperabilità, open data e privacy" della Direzione regionale per "l'Innovazione Tecnologica e la Trasformazione Digitale" all' Ing. Andrea Corbelli;

VISTO il Decreto Legislativo 30 giugno 2003, n.196 (Codice in materia di protezione dei dati personali);

VISTO il Decreto Legislativo 7 marzo 2005, n. 82 recante "Codice dell'amministrazione digitale", di seguito CAD;

VISTO l'art. 17 del suddetto CAD rubricato "Responsabile per la transizione digitale e difensore civico digitale", come di recente modificato, che prevede al comma 1 "…ciascuna pubblica amministrazione affida a un unico ufficio dirigenziale generale, fermo restando il numero complessivo di tali uffici, la transizione alla modalità operativa digitale e i conseguenti processi di riorganizzazione finalizzati alla realizzazione di un'amministrazione digitale e aperta, di servizi facilmente utilizzabili e di qualità, attraverso una maggiore efficienza ed economicità";

VISTO il Decreto Legislativo 195/2005 "Attuazione della direttiva 2003/4/CE sull'accesso del pubblico all'informazione ambientale" s.m.i.;

VISTO il Decreto Legislativo n. 36/2006 "Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico" relativo all'accesso all'informazione nel settore pubblico, attuativo della Direttiva 2003/98 relativa alla "public sector information" e s.m.i.;

VISTO il Decreto Legislativo 32/2010 "Attuazione della direttiva 2007/2/CE, che istituisce un'infrastruttura per l'informazione territoriale nella Comunità europea – INSPIRE" e s.m.i.;

CONSIDERATO il Provvedimento n. 243, Garante per la protezione dei dati personali, Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi,

Copia

effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati;

VISTO il Decreto Legislativo 14 marzo 2013, n. 33 (Riordino della disciplina riguardante il diritto di accesso civico e gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni);

CONSIDERATO il predetto D. Lgs. 33/2013, c.d. "Decreto Trasparenza", che statuisce norme che rendono obbligatoria la pubblicazione di documenti e dati con obbligo di pubblicità mediante la sezione "Amministrazione trasparente", la quale è luogo deputato all'accesso di tali materiali in formato aperto ai sensi dell'art. 68, comma 3, lettera a) del Codice dell'Amministrazione Digitale;

VISTO il D. Lgs. 97/2016, che prevede modifiche al c.d. "Decreto Trasparenza", d.lgs. 33/2013, introducendo il diritto d'accesso all'informazione pubblica;

CONSIDERATA la Legge 7 agosto 2015, n. 124 (Deleghe al Governo in materia di riorganizzazione delle amministrazioni pubbliche), che, all'art. 1, rubricato "Carta della cittadinanza digitale", afferma la necessità di "garantire ai cittadini e alle imprese, anche attraverso l'utilizzo delle tecnologie dell'informazione e della comunicazione, il diritto di accedere a tutti i dati, i documenti e i servizi di loro interesse in modalità digitale";

VISTA la Direttiva (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;

CONSIDERATE le Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico del 13 febbraio 2020, predisposte dall'Agenzia per l'Italia Digitale (AgID) con l'obiettivo di supportare le Amministrazioni pubbliche nel processo di valorizzazione del proprio patrimonio;

VISTO il Decreto Legislativo 8 novembre 2021, n. 200 (Attuazione della direttiva (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 (rifusione);

VISTO il Piano Triennale per l'informatica nella Pubblica Amministrazione per il triennio 2024-2026 di AgID, approvato con DPCM del 12 gennaio 2024, che intende promuovere la trasformazione digitale del Paese e, in particolare, della Pubblica Amministrazione, delineando le linee guida attraverso l'individuazione di principi, obiettivi e risultati attesi sostenibili e misurabili;

CONSIDERATA la Deliberazione del 24 maggio 2022, n. 326 che approva l'Agenda Digitale Lazio 2022-2026, individuando, altresì, come una delle sfide principali per le Pubbliche Amministrazioni quella degli "Open data", non solo come strumento indispensabile per un'Amministrazione trasparente ma anche come un'opportunità per farsi "promotrice dello sviluppo a livello territoriale", ponendo a disposizione della collettività il patrimonio di banche dati di proprietà dell'Amministrazione regionale, in coerenza con il percorso avviato anche a livello centrale;

Copia

RITENUTO che gli adempimenti derivanti dalla normativa vigente in materia di open data possano rappresentare, per l'amministrazione regionale, una opportunità per:

- rendere trasparenti e visibili le modalità operative di gestione dei dati a disposizione dell'amministrazione;

- contribuire al miglioramento della qualità della vita di tutti i cittadini che possono utilizzare i dati resi loro disponibili;

- contribuire alla condivisione di cataloghi open data sovraregionali.

VISTE le allegate Linee Guida redatte dall'Area "Sistemi informativi trasversali, interoperabilità, open data e privacy";

VISTA l'importanza della partecipazione attiva di tutti gli stakeholder interessati ai processi decisionali per favorire una maggiore inclusione e rappresentatività;

CONSIDERATA la disponibilità della piattaforma EUSURVEY, messa a disposizione gratuitamente dalla Commissione Europea, come strumento idoneo per agevolare la raccolta e l'analisi dei dati relativi alla consultazione in oggetto;

VISTA la Determinazione Dirigenziale n. G09460 del 16 luglio 2024 concernente "Approvazione e avvio consultazione pubblica delle "Linee Guida Regionali Open Data", con cui sono state approvate le Linee Guida Regionali Open Data ed è stato dato inizio alla consultazione delle suddette Linee Guida per un congruo un periodo di 30 giorni, attraverso i canali istituzionali della Regione Lazio;

PRESO ATTO che il suindicato periodo di 30 giorni, necessario a recepire i contributi ritenuti utili da parte dell'amministrazione regionale, è giunto a scadenza il 12 agosto 2024 senza che siano pervenuti contributi o osservazioni;

CONSIDERATO che il presente atto non comporta oneri di spesa in capo alla Regione Lazio;

RITENUTO quindi, alla luce di quanto esposto, di dover procedere all'approvazione definitiva delle allegate Linee Guida Regionali Open Data, che costituiscono parte integrante e sostanziale della presente determinazione;

DETERMINA

- di approvare le Linee Guida Regionali Open Data allegate, che costituiscono parte integrante e sostanziale della presente determinazione;

Copia

La presente determinazione dirigenziale con il relativo allegato sarà pubblicata sul Bollettino Ufficiale della Regione Lazio e sul sito Internet istituzionale della Regione Lazio www.regione.lazio.it, come previsto dal Decreto Legislativo n. 33/2013 “Riordino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle Pubbliche Amministrazioni”.

Il Direttore Regionale
Luca Marta

Copia

REGIONE LAZIO

LINEE GUIDA REGIONALI OPEN DATA

Copia

REGIONE LAZIO

INDICE DELLE FIGURE **4**

REVISIONI **5**

PREMESSA **6**

1. NORMATIVA E DOCUMENTI DI RIFERIMENTO **6**

2. PRINCIPI GENERALI **8**

2.1. DEFINIZIONI DI DATO E DATASET "APERTO" 8

2.2. PRINCIPI FAIR 9

2.3. DEFINIZIONE DI METADATI 10

2.4. FORMATI PIÙ COMUNI PER I DATI APERTI 11

2.5. SPECIFICHE CATEGORIE DI DATI APERTI (D. LGS. 24/01/2006, N. 36) 12

2.5.1. I dati dinamici *13*

2.5.2. Le serie di dati di "elevato valore" *13*

2.5.3. I dati della ricerca *15*

2.5.4. I dati territoriali o dati geospaziali *15*

2.6. VINCOLI GIURIDICI DI PUBBLICAZIONE 16

2.7. SCHEDA RIEPILOGATIVA 18

3. PROCESSO DI PUBBLICAZIONE **21**

3.1. DEFINIZIONE DEI RUOLI E DELLE RESPONSABILITÀ 23

3.2. IDENTIFICAZIONE DEI DATASET 26

3.3. ANALISI 30

3.3.1. Analisi della qualità *30*

3.3.2. Bonifica *31*

3.3.3. Analisi di processo *32*

3.4. ARRICCHIMENTO 34

3.5. MODELLAZIONE E DOCUMENTAZIONE 36

3.6. VALIDAZIONE 38

3.7. PUBBLICAZIONE 40

Copia

REGIONE LAZIO

3.7.1. Metadatazione e uso delle API 40
3.7.2. Aspetti architetturali relativi alla pubblicazione dei dataset aperti 41
3.7.3. Uso di identificatori univoci e persistenti 43
3.7.4. Politiche di accesso 44
3.8. COINVOLGIMENTO DELL'UTENZA FINALE 45
4. BUONE PRATICHE 47
4.1. DEFINIZIONE "NOTE" DEL TRACCIATO RECORD 47
4.2. CODIFICA DEI CARATTERI 47
4.3. VALORI SCONOSCIUTI O NULLI 47
4.4. SCELTA DEI NOMI DEI CAMPI 48
4.5. RAPPRESENTAZIONE WIDE E LONG 48
4.6. SUBTOTALI, TOTALI O RAGGRUPPAMENTI 49
4.7. TIPI DI DATI 49
4.8. STANDARDIZZAZIONE DEI VALORI DEI CAMPI 50
5. PANIERE DEI DATI 52
6. LICENZE 55
7. CONCLUSIONI 58
8. APPENDICE 59
8.1. METADATI: PRINCIPALI OBBLIGATORIETÀ E RACCOMANDAZIONI 59
8.2. MISURE PER LA VALUTAZIONE QUALITÀ STANDARD ISO/IEC 25024 61
8.3. CHECK-LIST VERIFICA LIMITI GIURIDICI 63
8.4. RACCOMANDAZIONI SULLA QUALITÀ DEI DATI 66
8.5. GLOSSARIO 67

Copia

REGIONE LAZIO

Indice delle figure

Figura 1 - PRINCIPALI FORMATI PER GLI OPEN DATA..... 12

Figura 2 - Il percorso di preparazione dei dati delle linee guida di AgID..... 21

Copia

REGIONE LAZIO

Revisioni

Versione	Data	Tipologia modifica
1.0	13/08/2024	Versione iniziale

Copia

REGIONE LAZIO

Premessa

Le "Linee Guida Regionali Open Data" della Regione Lazio, elaborate in applicazione delle "Linee Guida recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico"[1] di AgID, forniscono un quadro chiaro sulle modalità di gestione del processo di pubblicazione e aggiornamento dei dati aperti da parte delle Strutture regionali, promuovendo la trasparenza e il riutilizzo delle informazioni pubbliche.

La pubblicazione di dati aperti favorisce anche l'innovazione, la collaborazione e lo sviluppo di nuovi servizi e applicazioni. È importante valutare la qualità dei dati, assicurarsi che siano aggiornati e garantire la loro accessibilità e fruibilità.

È fondamentale approntare un percorso di formazione costante che consenta di approfondire le conoscenze e di adottare le migliori pratiche nel campo dell'apertura dei dati.

Copia

1.Normativa e documenti di riferimento

[CAD] Decreto legislativo 7 marzo 2005, n. 82 recante "Codice dell'amministrazione digitale"

[D-LGS-36-2006] Decreto legislativo 24 gennaio 2006, n. 36 recante "Attuazione della direttiva (UE) 2019/1024 relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico che ha abrogato la direttiva 2003/98/CE"

[INSPIRE-DIR] Direttiva 2007/2/CE del Parlamento europeo e del Consiglio, del 14 marzo 2007, che istituisce un'Infrastruttura per l'informazione territoriale nella Comunità europea (INSPIRE).

[1] https://www.agid.gov.it/sites/default/files/repository_files/lg-open-data_v.1.0_1.pdf

REGIONE LAZIO

[D-LGS-32-2010] Decreto legislativo 27 gennaio 2010, n. 32 recante "Attuazione della direttiva 2007/2/CE, che istituisce un'infrastruttura per l'informazione territoriale nella Comunità europea (INSPIRE)"

[D-LGS-33-2013] Decreto Legislativo 14 marzo 2013, n. 33 recante "Riordino della disciplina riguardante il diritto di accesso civico e gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni"

[D-LGS-102-2015] Decreto Legislativo 18 maggio 2015, n. 102 recante "Attuazione della direttiva 2013/37/UE che modifica la direttiva 2003/98/CE, relativa al riutilizzo dell'informazione del settore pubblico"

[OD-DIR] Direttiva (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 (rifusione).

[D-LGS-200-2021] Decreto Legislativo 8 novembre 2021, n. 200 recante "Attuazione della direttiva (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 (rifusione)"

[LG-OD] Linee Guida AgID recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico.

[REG-UE-HVD] Regolamento di esecuzione (UE) 2023/138 della Commissione del 21 dicembre 2022 che stabilisce un elenco di specifiche serie di dati di elevato valore e le relative modalità di pubblicazione e riutilizzo.

[PIANO-TRIENNALE] Piano Triennale per l'Informatica nella Pubblica Amministrazione, aggiornamento 2022-2024.

Copia

REGIONE LAZIO

2. Principi generali

Nel presente capitolo vengono chiariti alcuni concetti che definiscono le caratteristiche e l'ambito di applicazione dei cosiddetti dati aperti.

2.1. Definizioni di dato e dataset "aperto"

Nelle "Linee Guida recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico" – da qui in poi "Linee Guida Open Data nazionali" – si sottolinea come nelle normative correnti l'unica differenza tra "dato" e "documento" sia che il primo è la rappresentazione simbolica o numerica, in formato digitale, di un fatto, un concetto o un'idea, mentre il secondo svolge la funzione di contenitore, fisico o digitale, di tali dati, contenente dati il quale può servirsi di testi, immagini, tabelle, grafici, ecc.

I dati sono considerati aperti quando sono resi **disponibili gratuitamente** o a un costo marginale, pubblicati in **formati documentati** e **neutri tecnologicamente, leggibili meccanicamente** e **disaggregati**, associati a **licenze** e **metadati standard** (vedi Linee Guida Open Data nazionali[2]), in modo che possano essere **utilizzati**, **riutilizzati** e **condivisi liberamente** da chiunque e **per qualsiasi finalità**, nei limiti previsti dalla disciplina vigente.[3]

I requisiti principali dei dati aperti sono quindi:

- **Disponibilità e accesso:** deve essere possibile accedere ai dati senza restrizioni non riconducibili a esplicite norme di legge, gratuitamente o a un prezzo non superiore ad un ragionevole costo di riproduzione, preferibilmente mediante procedura di download da un sito web pubblico. I dati devono essere disponibili in un formato strutturato che consenta alle applicazioni software di individuare, riconoscere ed estrarre facilmente dati specifici.

[2] https://www.agid.gov.it/sites/default/files/repository_files/lg-open-data_v.1.0_1.pdf
[3] https://opendatahandbook.org/guide/it/what-is-open-data/

- **Riutilizzo e ridistribuzione:** i dati devono essere forniti a condizioni tali da permetterne il riutilizzo e la ridistribuzione.
- **Partecipazione universale:** al fine di abbattere ogni barriera o discriminazione, chiunque deve avere uguali opportunità di accedere, utilizzare e contribuire ai dati aperti del settore pubblico, come **bene comune digitale**.

NOTA BENE: nel contesto dei dati aperti, i termini "**dati aperti**" e "**dataset aperto**" vengono frequentemente utilizzati come se fossero intercambiabili. Tuttavia, è importante riconoscere che essi rappresentano in realtà concetti differenti. Il dataset è infatti una raccolta specifica e organizzata di dati aperti.

2.2. Principi FAIR

Nelle Linee Guida Open Data nazionali si **raccomanda** di seguire e applicare i **principi del *framework FAIR***, vale a dire **reperibilità**, **accessibilità**, **interoperabilità** e **riutilizzabilità**.

Si tratta di una raccomandazione con diversi punti di sovrapposizione con i **requisiti generali** elencati nelle Linee Guida Open Data nazionali, a cui si rimanda per un approfondimento per comprendere nel dettaglio le differenze.

A seguire, alcuni punti di interesse dei principi *FAIR*, che qui si raccomanda di applicare in modo generale alla pubblicazione dei dati della Regione Lazio:

- per rendere i dati **interoperabili**, è necessario:
  - utilizzare, ove possibile, vocabolari controllati, parole chiave, *thesauri* e ontologie;
  - inserire riferimenti qualificati ad altri dati o metadati;
- per rendere i dati **riusabili**, è necessario:
  - rendere disponibili dati accurati e ben descritti con molteplici attributi pertinenti;
  - rendere chiaro come, perché, quando e da chi i dati sono stati creati e processati.

Copia

REGIONE LAZIO

2.3. Definizione di metadati

Con il termine "**metadati"** ci si riferisce all'insieme delle informazioni strutturate che costituiscono i dati che descrivono altri dati [4]. Rappresentano, quindi, le informazioni relative ai dati a cui si riferiscono e sono legati al loro ciclo di vita. I metadati consentono una maggiore comprensione dei dati di riferimento e rappresentano la chiave attraverso cui abilitare più agevolmente la ricerca, la scoperta, l'accesso e quindi il riuso dei dati.

Nell'ambito delle attività istituzionali AgID finalizzate alla promozione delle politiche di valorizzazione del patrimonio informativo pubblico nazionale e all'attuazione della Direttiva relativa al riutilizzo dei dati pubblici (nota come Direttiva PSI 2.0[5]), è stato elaborato il profilo italiano dei metadati richiesti per descrivere tutti i dati disponibili presso la Pubblica Amministrazione. Il profilo nazionale, che si inserisce nel contesto del *framework* europeo di interoperabilità[6], rappresenta un'estensione della specifica DCAT-AP[7] (*Data Catalog Vocabulary – Application Profile*), definita nell'ambito del programma ISA[8] (*Interoperability Solutions for European Public Administrations*) della Commissione Europea. Per tale motivo, questo è stato denominato anche "**Profilo italiano di DCAT-AP**[9]" – abbreviato in "*DCAT-AP_IT*". Il profilo applicativo europeo, così come la raccomandazione DCAT, si pone come obiettivo quello di proporre un modello per la rappresentazione di metadati indipendenti dal particolare dominio applicativo.

Per i metadati descrittivi generali relativi ai dati aperti, ovvero non dipendenti dalle tipologie di dati, deve quindi essere applicato il **profilo nazionale DCAT-AP_IT**,

[4] Un **metadato** (dal greco antico μετὰ, «oltre, dopo, per mezzo» e dal latino *datum*, «informazione» - plurale: *data*), letteralmente "(dato) per mezzo di un (altro) dato", è un'informazione che descrive un insieme di dati.
[5] https://eur-lex.europa.eu/eli/dir/2019/1024/oj/ita
[6] https://eur-lex.europa.eu/legal-content/IT/TXT/?qid=1584086617794&uri=CELEX:52017DC0134
[7] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/11
[8] https://ec.europa.eu/isa2/home_en/
[9] https://www.dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0

Copia

REGIONE LAZIO

rispettando le obbligatorietà, le raccomandazioni e seguendo gli esempi così come definiti nella relativa specifica e ontologia (vedi paragrafo 8.1).

2.4. Formati più comuni per i dati aperti

Per la scelta del formato dei dati aperti, un contributo utile si può rinvenire nella **Errore. L'origine riferimento non è stata trovata.**, che elenca alcuni tra i formati più utilizzati, specificando quali sono leggibili meccanicamente e non proprietari. Per misurare il livello di qualità dei formati più idonei per la pubblicazione, si attribuisce un punteggio che adotta come simbolo una stella. I livelli di qualità vanno da un minimo di una a un massimo di cinque stelle e tale misurazione compare nella tabella come ultima colonna.

I formati evidenziati in verde sono quelli preferenziali; se ciò non è possibile allora si possono utilizzare quelli evidenziati in giallo, mentre sono da escludere quelli evidenziati in rosso. In linea generale, è opportuno utilizzare il formato più adatto al tipo di dato che si vuole pubblicare e adottare formati con almeno **tre stelle**, in quanto rispondono in modo adeguato ai requisiti che i dati aperti devono soddisfare, soprattutto per quanto riguarda la leggibilità meccanica e il fatto di non appartenere a un formato proprietario.

Copia

REGIONE LAZIO

Format	Non-proprietary	Machine readable	Achievable stars
RDF	Yes	Yes	✪ ✪ ✪ ✪
XML	Yes	Yes	✪ ✪ ✪
JSON	Yes	Yes	✪ ✪ ✪
CSV	Yes	Yes	✪ ✪ ✪
ODS	Yes	Predominantly	✪ ✪ ✪
XLSX	Yes	Predominantly	✪ ✪ ✪
XLS	No	Predominantly	✪ ✪
TXT	Yes	Predominantly	✪*
HTML	Yes	Predominantly	✪*
PDF	Yes	No	✪
DOCX	Yes	No	✪
ODT	Yes	No	✪
PNG	Yes	No	✪
GIF	No	No	✪
JPG/JPEG	No	No	✪
TIFF	No	No	✪
DOC	No	No	✪

FIGURA 1 - PRINCIPALI FORMATI PER GLI OPEN DATA

2.5. Specifiche categorie di dati aperti (D. Lgs. 24/01/2006, n. 36)

Per quanto riguarda le categorie di dati aperti descritte nelle Linee Guida Open Data nazionali, le quali recepiscono quanto stabilito dal D. Lgs. del 24/01/2006 n. 36, di particolare interesse sono: i "**dati dinamici**", i dati "**di elevato valore**", i dati "**della ricerca**" e i dati "**territoriali o geospaziali**".

REGIONE LAZIO

2.5.1. I dati dinamici

I dati dinamici si riferiscono alle informazioni che, a causa della natura volatile e della rapida obsolescenza che le caratterizza, sono soggetti ad aggiornamenti frequenti o in tempo reale.

Appartengono a questa tipologia i dati generati da sensori, ambientali, relativi al traffico, satellitari o meteorologici.

I dati dinamici sono una delle tipologie più preziose dal punto di vista commerciale, in quanto possono essere utilizzati per prodotti e servizi che forniscono informazioni in tempo reale, come app di viaggio o di trasporto. Considerato che il valore economico di tale tipologia di dati dipende dall'immediata disponibilità dell'informazione e dagli aggiornamenti regolari, come indicato nel Decreto sopra citato, i dati dinamici, oltre ad essere coerenti con i principi generali relativi ai dataset aperti, devono essere resi disponibili per il riutilizzo immediatamente dopo la raccolta[10] tramite API[11] adeguate e, ove possibile, attraverso il download massivo.

2.5.2. Le serie di dati di "elevato valore"

I dati di "elevato valore" vengono individuati[12] per la loro capacità di:

- apportare importanti benefici, socioeconomici o ambientali, e contribuire all'offerta di servizi innovativi;

Copia

---

[10] Se, per motivi di capacità finanziarie o tecniche, ciò non sia possibile, allora i dati dinamici possono essere resi disponibili per il riutilizzo entro un termine e/o con temporanee restrizioni tecniche, tali, però, da non pregiudicare lo sfruttamento del loro potenziale economico e sociale. Il termine e le restrizioni tecniche di cui sopra devono essere definiti e motivati con apposito provvedimento del titolare dei dati.

[11] Con il termine API (Application Programming Interface) si intende un'interfaccia per programmi applicativi costituita da un insieme di funzioni, procedure, definizioni e protocolli per la comunicazione da macchina a macchina e lo scambio ininterrotto di dati. Le API sviluppate devono essere conformi alle "Linee Guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni" e le "Linee Guida Tecnologie e standard per la sicurezza dell'interoperabilità tramite API dei sistemi informatici", adottate con la Determinazione di AgID n. 547/2021.

[12] Art. 14 c 12 Direttiva (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 (rifusione)

REGIONE LAZIO

- beneficiare un numero elevato di utilizzatori, in particolare Piccole-Medio Imprese (PMI);
- contribuire a generare proventi;
- essere combinati con altre serie di dati.

Le serie di dati di elevato valore, individuate dal Regolamento di esecuzione 2023/138 del 21 dicembre 2022 della Commissione Europea[13], devono essere messe a disposizione gratuitamente[14] e rese disponibili attraverso il download massivo e, come per i dati dinamici, tramite API.

Le serie di dati di elevato valore individuate dalla Commissione fanno riferimento alle categorie tematiche indicate nell'Allegato I della Direttiva (UE) 2019/1024, vale a dire:

1. Dati geospaziali
2. Dati relativi all'osservazione della terra e all'ambiente
3. Dati meteorologici
4. Dati statistici
5. Dati relativi alle imprese e alla proprietà delle imprese
6. Dati relativi alla mobilità

Nell'Allegato del suddetto regolamento sono dettagliate le serie di dati di elevato valore, per ciascuna di queste categorie.

Il regolamento si applicherà a decorrere da 16 mesi dopo la sua entrata in vigore, ovvero dal 9 giugno 2024.

Copia

[13] *https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32023R0138*

[14] *Fanno eccezione i dati di elevato valore che rientrano nelle seguenti casistiche:*

- *in possesso delle imprese pubbliche qualora ciò determini una distorsione della concorrenza nei pertinenti mercati, se così previsto dagli atti di esecuzione di cui sopra;*
- *in possesso di biblioteche, comprese le biblioteche universitarie, musei e archivi;*
- *che, in possesso degli enti pubblici, generano utili per coprire una parte sostanziale dei costi inerenti allo svolgimento dei compiti di servizio pubblico e la loro messa a disposizione gratuita avrebbe un impatto sostanziale sul bilancio di tali enti. In tal caso la possibile esenzione dall'obbligo di mettere a disposizione gratuitamente le serie di dati di valore elevato deve durare per un periodo non superiore ai due anni dall'entrata in vigore del pertinente atto di esecuzione della Commissione Europea.*

REGIONE LAZIO

2.5.3. I dati della ricerca

I dati della ricerca sono definiti come *"documenti informatici, diversi dalle pubblicazioni scientifiche, raccolti o prodotti nel corso della ricerca scientifica e utilizzati come elementi di prova nel processo di ricerca, o comunemente accettati nella comunità di ricerca come necessari per convalidare le conclusioni e i risultati della ricerca"*.

Esempi di tale tipologia di dati sono: statistiche, risultati di esperimenti, misurazioni, osservazioni risultanti dall'indagine sul campo, risultati di indagini, immagini e registrazioni di interviste, oltre a metadati, specifiche e altri oggetti digitali.

I dati da considerare sono quelli che rappresentano il risultato di attività di ricerca finanziata con fondi pubblici e resi pubblici, anche attraverso l'archiviazione in una banca dati pubblica, gestita a livello istituzionale o su base tematica, da ricercatori, organizzazioni che svolgono attività di ricerca e organizzazioni che finanziano la ricerca.

Nel processo di identificazione dei dati della ricerca da rendere disponibili per il riutilizzo è necessario tenere conto della protezione dei dati personali, degli interessi commerciali, dei diritti di proprietà intellettuale[15] e dei diritti di proprietà industriale[16].

2.5.4. I dati territoriali o dati geospaziali

I dati territoriali sono definiti dal CAD come "*i dati che attengono, direttamente o indirettamente, a una località o a un'area geografica specifica*" (cfr. art. 1 c. 1 lettera i-sexies), in linea con l'analoga definizione presente all'art. 3 punto 2 della Direttiva [INSPIRE-DIR].

Per le disposizioni su tali dati si rimanda alla disciplina specifica definita in applicazione della Direttiva 2007/2/CE conosciuta come Direttiva INSPIRE [INSPIRE-DIR] e

[15] *v. legge 22 aprile 1941, n. 633*

[16] *v. decreto legislativo 10 febbraio 2005, n. 30*

REGIONE LAZIO

recepita in Italia con il decreto legislativo 27 gennaio 2010, n. 32 [D-LGS-32-2010] o ai requisiti specifici per la categoria "Dati geospaziali" prevista nell'ambito dei dati di elevato valore.

Nel caso di dati territoriali, quindi, si deve fare riferimento ai Regolamenti europei e alle norme nazionali e relativi documenti tecnici per le diverse componenti (metadati, dati e servizi di dati, servizi di rete, condivisione e monitoraggio) dell'infrastruttura istituita e implementata nell'ambito di INSPIRE.

Sono fatte salve tutte le altre norme applicabili ai dati territoriali, come la legge n. 132/2016[17] relativamente al Sistema Informativo Nazionale Ambientale (SINA) e alla rete SINANET.

Per l'apertura e il riutilizzo dei dati territoriali si applicano le indicazioni e i requisiti generali, cioè validi per tutte le tipologie di dati, definiti nelle Linee Guida Open Data nazionali, facendo riferimento, per i formati, a quelli specifici per tale tipologia di dati.

Nel caso in cui i dati territoriali siano anche dati dinamici oppure dati della ricerca, o invece, rientrino tra le serie di dati di elevato valore, allora devono essere applicati anche i requisiti specifici di quella categoria di dati e, nel caso dei dati di elevato valore, si devono applicare le indicazioni presenti negli atti di esecuzione adottati dalla Commissione Europea per la categoria "Dati geospaziali".

2.6. Vincoli giuridici di pubblicazione

I principali vincoli per la pubblicazione dei dati sono correlati alla protezione dei dati personali.

Vi è l'esclusione dall'ambito di applicazione della norma sui dati aperti di documenti che contengono dati personali il cui riuso è incompatibile con gli scopi originari del

[17] *https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:2016;132%7Eart2*

REGIONE LAZIO

trattamento da parte della Pubblica Amministrazione. Ciò implica che i dati personali, di fatto, non possono essere considerati open data a meno che esista una norma che preveda espressamente la facoltà o l'obbligo di pubblicarli.

Per supportare l'analisi dei vincoli sui dati personali l'AgID ha inserito nelle Linee Guida Open Data nazionali una check-list, riportata in appendice, d'ausilio alle pubbliche amministrazioni, per verificare gli aspetti giuridici che potrebbero incidere sulla possibilità di considerare "open" alcuni dataset.

L'obiettivo è quello di verificare che non siano resi pubblici dati che potrebbero essere ricondotti a persone fisiche[18] e che, quindi, potrebbero ledere i loro diritti e le loro libertà.

Tra gli altri aspetti della check-list, vi sono: la verifica della proprietà intellettuale della sorgente dati; la verifica della segretezza di dati che, se resi pubblici, potrebbero comportare un pericolo per la sicurezza nazionale; la verifica di norme di trasparenza che impediscano, ad esempio, l'indicizzazione da parte dei motori di ricerca.

In ultimo, si sottolinea come nell'ambito della Strategia Cloud Italia sia stata definita una classificazione dei dati e dei servizi[19] , **strategici**, **critici** e **ordinari,** allo scopo di definire un processo di classificazione dei dati in base al danno che una loro compromissione potrebbe provocare al Paese.).

Copia

[18] *Occorre valutare con attenzione i dati inerenti a:*

- *libertà e segretezza della corrispondenza;*
- *interessi economici e commerciali di una persona fisica o giuridica, ivi compresi la proprietà intellettuale, il diritto d'autore e i segreti commerciali;*
- *dati amministrativi contenenti informazioni di carattere psicoattitudinale relativi a terzi;*

[19] *https://docs.italia.it/italia/cloud-italia/strategia-cloud-italia-docs/it/stabile/4_la_strategia_cloud_per_la_pubblica_amministrazione.html#la-classificazione-dei-dati-e-dei-servizi*

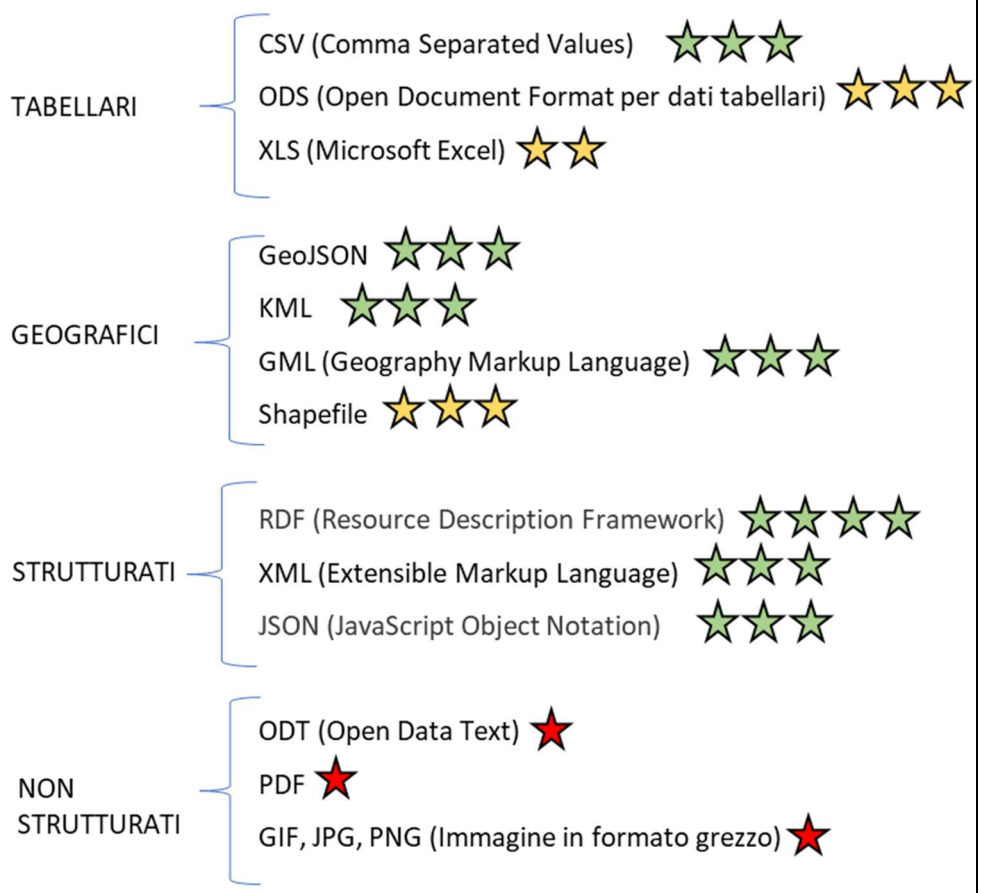
REGIONE LAZIO

2.7. Scheda riepilogativa

Scheda Riepilogativa		
Definizioni		
Dati Aperti	Definizione	I dati aperti sono dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque, eventualmente citando la fonte e/o condividendoli con lo stesso tipo di licenza di rilascio. I dati aperti devono quindi favorire il riuso da parte di altri soggetti pubblici o privati, per essere utilizzati in ambiti differenti da quelli per i quali sono stati prodotti o raccolti.
	Requisiti	Requisito giuridico: Licenza aperta. I dati devono essere rilasciati e distribuiti con una Licenza di tipo aperto, che non impedisca, anzi incentivi il riuso dei dati anche a fini commerciali. La licenza può contenere alcuni diritti riservati, per esempio la richiesta di citazione della fonte dei dati
		Requisito economico: Gratuiti I dati devono essere disponibili al riuso gratuitamente o, per specifici ed eccezionali casi, a costi marginali
		Requisito tecnologico: Formati aperti Dati rilasciati in formati aperti, non vincolati tecnicamente, ossia con una specifica tecnica di pubblico dominio e che rispettino standard e specifiche che rendano i dati facilmente leggibili da macchina, utilizzabili in automatico dal computer e da applicazioni informatiche (machine-readable)
Dataset	Definizione	Un dataset è una collezione organizzata di dati omogenei generalmente riguardanti una stessa organizzazione o categoria
Metadati	Definizione	I metadati rappresentano la chiave attraverso cui abilitare più agevolmente la ricerca, la scoperta, l'accesso e quindi il riuso dei dati stessi.
	Requisito	I dati, con esclusione di quelli territoriali, resi disponibili per il riutilizzo devono essere documentati attraverso metadati conformi al profilo DCAT-AP_IT definito con la guida operativa per i cataloghi dati.
Formati	Indicazioni	Il 'formato' si riferisce al modo in cui i dati sono strutturati e resi disponibili per l'uso da parte di persone e di macchine. Si devono selezionare i formati che meglio si adattano al contenuto e ai dati da condividere e rilasciare

Copia

REGIONE LAZIO

	Esempi di formato da usare in base al tipo di dato da pubblicare	
Dati oggetto di eventuale pubblicazione		
Dati dinamici	Definizione	Si tratta dei documenti informatici, soggetti ad aggiornamenti frequenti o in tempo reale, in particolare a causa della loro volatilità o rapida obsolescenza. Appartengono a questa tipologia i dati generati da sensori, ma anche i dati ambientali, relativi al traffico, satellitari o meteorologici
Dati di elevato valore	Definizione	Sono quei "documenti il cui riutilizzo è associato a importanti benefici per la società, l'ambiente e l'economia, in considerazione della loro idoneità per la creazione di servizi, applicazioni a valore aggiunto e nuovi posti di lavoro, nonché del numero dei potenziali beneficiari dei servizi e delle applicazioni a valore aggiunto basati su tali serie di dati
	Criteri	• Per l'identificazione di tale tipologia di dati deve essere valutata la loro potenzialità: nell'apportare importanti benefici socioeconomici o ambientali e servizi innovativi; • nel beneficiare un numero elevato di utilizzatori, in particolare PMI; • nel contribuire a generare proventi; • nell'essere combinati con altre serie di dati.

REGIONE LAZIO

	Limitazioni	Le serie di dati di elevato valore, individuate da specifici atti di esecuzione della Commissione Europea, devono essere messe a disposizione gratuitamente ad eccezione di quelle: • in possesso delle imprese pubbliche qualora ciò determini una distorsione della concorrenza nei pertinenti mercati, se così previsto dagli atti di esecuzione di cui sopra; • in possesso di biblioteche, comprese le biblioteche universitarie, musei e archivi; • che, in possesso degli enti pubblici, generano utili per coprire una parte sostanziale dei costi inerenti allo svolgimento dei compiti di servizio pubblico e la loro messa a disposizione gratuita avrebbe un impatto sostanziale sul bilancio di tali enti. In tal caso la possibile esenzione dall'obbligo di mettere a disposizione gratuitamente le serie di dati di valore elevato deve durare per un periodo non superiore ai due anni dall'entrata in vigore del pertinente atto di esecuzione della Commissione Europea.
	Categorie tematiche	Dati geospaziali
		Dati relativi all'osservazione della terra e all'ambiente
		Dati meteorologici
		Dati statistici
		Dati relativi alle imprese e alla proprietà delle imprese
		Dati relativi alla mobilità
Dati della ricerca	Definizione	Documenti informatici, diversi dalle pubblicazioni scientifiche, raccolti o prodotti nel corso della ricerca scientifica e utilizzati come elementi di prova nel processo di ricerca, o comunemente accettati nella comunità di ricerca come necessari per convalidare le conclusioni e i risultati della ricerca
	Requisiti	I dati come tutti i risultati della ricerca devono innanzitutto essere conformi ai criteri FAIR e cioè essere reperibili tempestivamente, accessibili su richiesta, interoperabili e comunque riusabili con le opportune regole, strumenti e risorse.
Dati territoriali (o geospaziali)	Definizione	Sono i dati che attengono, direttamente o indirettamente, a una località o a un'area geografica specifica

copia

REGIONE LAZIO

3. Processo di pubblicazione

Le linee Guida Open Data nazionali indicano un possibile percorso di preparazione dei dati, per garantirne la produzione e la pubblicazione di qualità, che può essere un riferimento per l'applicazione alle diverse realtà amministrative. Uno schema analogo, sebbene riferito ai *linked open data*, è definito anche nelle Linee Guida per l'interoperabilità semantica attraverso i *Linked Open Data*[20]. A tale proposito, si raccomanda di seguire le suddette Linee Guida per il processo di produzione di *Linked Open Data*.

Le fasi indicate nelle Linee Guida Open Data nazionali sono sei (vedi Figura 2):

1. Identificazione
2. Analisi
3. Arricchimento
4. Modellazione e Documentazione
5. Validazione
6. Pubblicazione

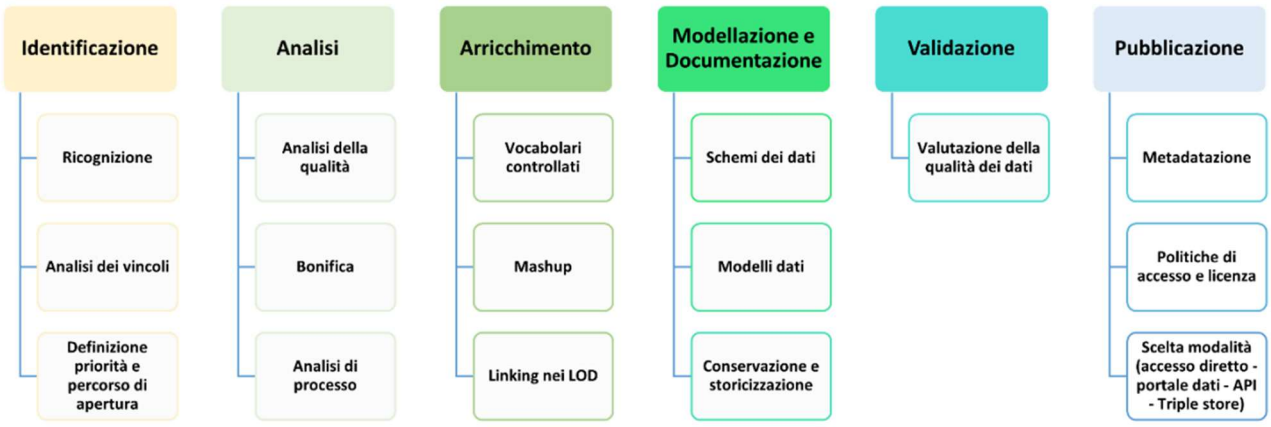


FIGURA 2 - IL PERCORSO DI PREPARAZIONE DEI DATI DELLE LINEE GUIDA OPEN DATA NAZIONALI

Alcune fasi indicate nel percorso possono non essere considerate, stante il possibile diverso punto di partenza che porterà alla pubblicazione dei dati. Se, per esempio, si

[20] *v.* https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf

REGIONE LAZIO

tratta di produrre un nuovo dato, quindi non ancora esistente, le attività di ricognizione e di analisi possono essere tralasciate e la prima fase da considerare potrà essere quella della modellazione, oltre alla definizione di priorità e del percorso di apertura, inserita nella fase di identificazione, che rimane comunque valida.

L'applicazione del processo, come descritto nelle Linee Guida Open Data nazionali (capitolo 5.1), deve avvenire in maniera costante: le attività non si esauriscono con la mera pubblicazione dei dati, ma devono prevedere un costante **aggiornamento, monitoraggio e coinvolgimento** dell'utenza finale, fasi non rappresentate nel percorso, ma ugualmente importanti per un'ulteriore ottimizzazione dei dati basata sulle effettive esigenze degli utenti.

Copia

REGIONE LAZIO

3.1. Definizione dei ruoli e delle responsabilità

Ai fini della corretta attuazione delle presenti Linee Guida, è definito il presente **modello di *governance* degli Open Data** al fine di garantire un solido presidio di tutte le fasi del processo di pubblicazione dei dati (par. 3 del presente documento).

Il modello organizzativo proposto si articola su due livelli:

- **Strategico**: definisce tutti gli aspetti di indirizzo relativi alla pubblicazione dei dati aperti, garantisce che la Strategia Open Data regionale sia costantemente aggiornata e allineata alla normativa comunitaria e nazionale di riferimento, verifica che le Linee Guida siano correttamente attuate e identifica i dataset da pubblicare (par. 3.2);
- **Operativo:** attua le presenti Linee Guida ed esegue le attività inerenti alla pubblicazione dei dati (da par. 3.3 a 3.8).

Di seguito si rappresenta la struttura organizzativa proposta:

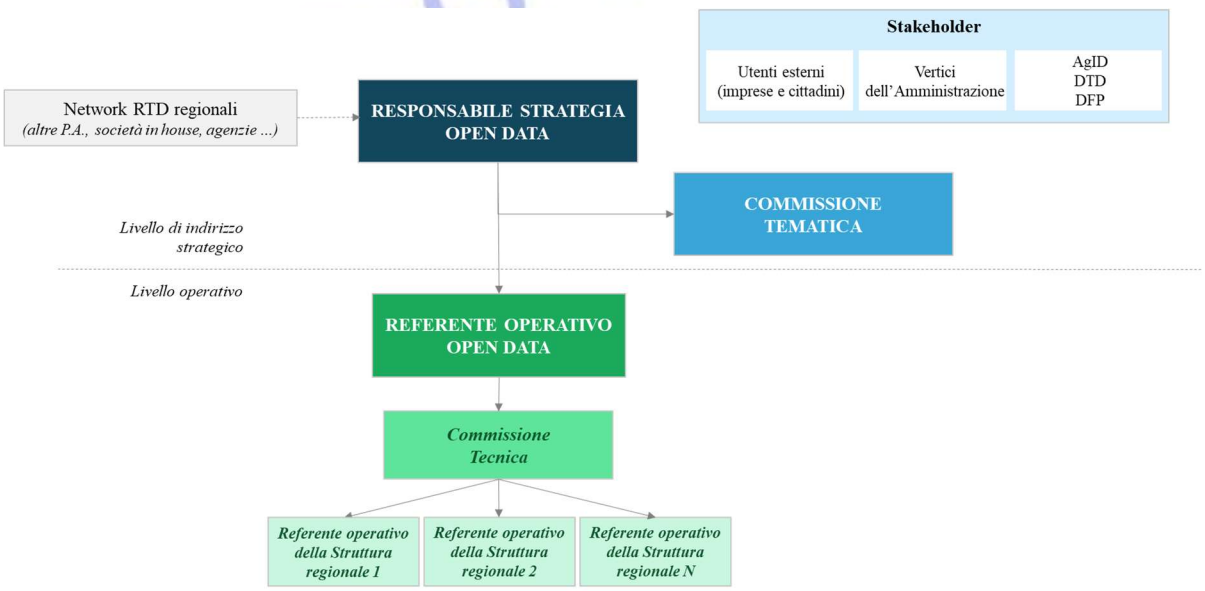


FIGURA 3 MODELLO DI GOVERNANCE

REGIONE LAZIO

Il governo degli Open Data è assicurato a livello di indirizzo strategico tramite le seguenti figure:

- Il ***Responsabile Strategia Open Data***, che in attuazione delle Linee Guida Regionali Open Data, coincide con il **Responsabile per la Transizione al Digitale (RTD)**, assolve il compito di (i) definire e aggiornare la strategia regionale per la gestione dei dati aperti, (ii) curare l'aggiornamento continuo delle Linee Guida e (iii) verificare la corretta applicazione delle Linee Guida all'interno dell'Amministrazione regionale;
- Gli ***Stakeholder*** possono essere consultati dal Responsabile Strategia Open data per la definizione delle linee di indirizzo strategico dell'Amministrazione. In particolare, i principali stakeholder sono: gli utenti esterni (imprese e cittadini); i vertici dell'Amministrazione; AgID, il DTD e il Dipartimento Funzione Pubblica;
- La ***Commissione Tematica*** è presieduta dal Responsabile Strategia Open Data o da suo delegato ed è composta dai Direttori delle Strutture regionali (o loro delegati) e dal Referente Operativo Open Data. Si riunisce di norma trimestralmente o su convocazione per specifiche esigenze per (i) identificare i dati che, ai sensi della normativa di riferimento, sono classificati come "dati di tipo aperto", (ii) effettuare una ricognizione dei dati già aperti e messi a disposizione dalla Regione e (iii) definire i dataset da pubblicare sul portale dedicato.

  La Commissione Tematica può consultare, ove necessario, il RPD regionale per la verifica della conformità dei dati da pubblicare alla normativa in materia di protezione dei dati personali.

Le attività di sviluppo della strategia Open Data sono realizzate a livello operativo tramite le seguenti figure:

- Il ***Referente Operativo Open Data***, individuato dal RTD, ha il compito di (i) governare e coordinare le attività connesse agli open data e al loro

Copia

REGIONE LAZIO

aggiornamento, (ii) applicare il modello regionale degli Open Data e monitorarne il corretto funzionamento e (ii) verificare la conformità delle prassi operative rispetto alla normativa di riferimento e alle Linee Guida.

Da un punto di vista operativo può avvalersi del supporto della Commissione Tematica, competente in materia di definizione dei dataset da pubblicare sul portale dati.lazio.it, della Commissione Tecnica, che ha il ruolo di supportarlo nella gestione operativa del modello. Inoltre, il Referente Operativo Open Data potrà essere eventualmente supportato nell'espletamento delle sue funzioni da una task-force dedicata. Per la condivisione di best practice di altre Amministrazioni, il Referente Operativo può consultarsi con gli RTD presenti sul territorio regionale;

- La ***Commissione Tecnica*** è presieduta e convocata dal Referente Operativo Open Data e composta dai referenti tecnici delle Strutture regionali, individuati dai Direttori, e dai referenti tecnici di LAZIOcrea.

  Si riunisce di norma trimestralmente o su convocazione per specifiche esigenze al fine di attuare le procedure operative per l'analisi della qualità del dato, la pubblicazione, la modellazione e l'aggiornamento dei dati pubblicati sul portale.

  La Commissione Tecnica si occupa di raccogliere le istanze dei referenti operativi al fine di adottare modalità operative uniformi.

  La Commissione Tecnica può avvalersi del Responsabile regionale per la gestione e la conservazione documentale per tutte le tematiche inerenti la conservazione del dato.
- I **referenti operativi delle Strutture regionali** attuano gli indirizzi delle seguenti Linee Guida nonché della Commissione Tematica e partecipano alla Commissione Tecnica.

Copia

REGIONE LAZIO

3.2. Identificazione dei dataset

La fase di identificazione dei dati, come indicato in **Errore. L'origine riferimento non è stata trovata.**, prevede tre sottofasi: ricognizione, analisi dei vincoli, definizione di priorità e percorso di apertura.

La ricognizione deve essere gestita dalla Commissione Tematica, e quindi dai Direttori (o loro delegati) di ogni Direzione e Agenzia regionale, che dovrà effettuare una fase di mappatura dei dati complessivamente disponibili nel proprio settore, eventualmente basandosi su mappature già fatte per altri scopi.

In primo luogo, si dovranno identificare le strutture funzionali all'interno della Direzione/Agenzia che detengono dati potenzialmente utili.

Dopo aver identificato le strutture funzionali, è necessario comprendere, nel loro ambito, quali sono le fonti dati di rilievo, cioè identificare archivi strutturati, elenchi, basi dati prodotte da software in uso presso l'amministrazione.

Per eseguire la mappatura si dovranno anche considerare le eventuali richieste provenienti dagli Stakeholder esterni alla Regione, come cittadini, sviluppatori, organizzazioni, associazioni, o altri portatori di interesse verso la pubblicazione di dati pubblici.

Una volta identificate le fonti dati, è necessario procedere ad una ricognizione dei dati o delle tipologie di dato effettivamente in possesso dell'Amministrazione. Tali dati possono essere:

- nativi, cioè dati generati dall'Amministrazione, perché prodotti dai vari uffici durante l'adempimento delle proprie funzioni istituzionali;
- mashup, cioè dati provenienti da diverse fonti e soggetti a operazioni di integrazione.

In questa fase sono da considerare i dati che le norme prevedono di rendere disponibili come dati di tipo aperto. In particolare, tra gli altri:

Copia

REGIONE LAZIO

- "i documenti, le informazioni e i dati oggetto di pubblicazione obbligatoria ai sensi della normativa vigente, resi disponibili anche a seguito dell'accesso civico"[21];
- tutti i dati acquisiti e generati nella fornitura dei servizi svolti in concessione da pubbliche amministrazioni[22]. L'obbligo del concessionario di rendere disponibili all'amministrazione concedente detti dati come dati di tipo aperto DEVE essere inserito nei contratti e nei capitolati con i quali le pubbliche amministrazioni affidano lo svolgimento dei servizi in concessione.

Si dovrà inoltre effettuare una ricognizione dei dati già aperti e messi a disposizione dalla Regione.

Come output dei passaggi precedentemente descritti si otterrà una prima mappa delle informazioni a disposizione dell'Amministrazione, che potrà essere sintetizzata in una tabella simile alla seguente.

Struttura Funzionale	Fonte dati	Tipologia di dato	Pubblicato

Nella fase di analisi dei vincoli, i dataset identificati vengono analizzati per comprendere se esistano ostacoli alla loro apertura. Si dovranno quindi verificare o risolvere eventuali impedimenti di riservatezza o titolarità del dato e controllare che il dataset sia correttamente reso in formato aperto, cioè che sia dotato di valida licenza per la pubblicazione e che sia in linea con la vigente normativa sulla protezione dei dati personali.

[21][D-LGS-33-2013],
[22] art. 50-quater del [CAD]

REGIONE LAZIO

Gli ostacoli più comuni alla pubblicazione di un dataset in formato aperto si riferiscono a:

- Proprietà dei dati
- Presenza di dati particolari
- Costi di estrazione dalla fonte originaria o di elaborazione non sostenibili
- Possibili conseguenze negative per la pubblica sicurezza
- Altri problemi di natura giuridica

Per supportare tale analisi, si potrà fare riferimento anche alla *check list* di verifica dei limiti giuridici presente in appendice.

Successivamente sarà necessario definire la priorità di apertura basandosi su fattori come l'impatto economico e sociale che deriverebbero dall'apertura dei dati, nonché del livello di interesse e delle necessità degli utilizzatori.

Si può quindi definire una scala di priorità basandosi sui seguenti criteri:

- i dati che le norme prevedono di rendere disponibili obbligatoriamente come dati di tipo aperto. Come quelli del "Paniere dinamico di dataset" di AgID (vedi capitolo "Paniere dei dati") e i dati di "elevato valore" definiti nell'allegato[23] del "Regolamento di esecuzione (UE) 2023/138", citati del capitolo "Le serie di dati di "elevato valore";
- i dati oggetto di richieste pervenute direttamente da soggetti interessati al riuso;
- le tipologie di dati a cui è riconosciuto un elevato potenziale economico e dati della ricerca;
- dati di potenziale interesse per la comunità (ad esempio dati che riguardano bilanci, inquinamento, epidemiologie sanitarie, ecc.).

[23] https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32023R0138#d1e32-48-1

REGIONE LAZIO

Tale attività viene sistematizzata attraverso la definizione di un percorso di apertura dei dati inserito all'interno del Piano Triennale ICT regionale.

Al fine di un'efficace rappresentazione dei ruoli e delle responsabilità, nel seguito si utilizza il metodo della matrice RACI[24].

FASE	R	A	C	I
Ricognizione	Referente operativo della Struttura regionale	Commissione Tematica	-	Commissione Tecnica
Analisi dei vincoli	Commissione Tematica	Commissione Tematica	Stakeholder	Commissione Tecnica
Definizione di priorità e percorso di apertura	Commissione Tematica	Commissione Tematica	Stakeholder	Commissione Tecnica

[24] Questa matrice è uno strumento visivo utilizzato per organizzare e comunicare chi è coinvolto in un progetto o in un'attività e quali sono le sue responsabilità.
Ogni cella della matrice corrisponde a una specifica combinazione di attività e ruolo. Le iniziali corrispondenti al ruolo vengono inserite all'interno di ciascuna cella per indicare chi è coinvolto e quali attività vengono svolte:

- R (Responsabile): La persona incaricata di eseguire l'attività.
- A (Accountable): La persona che è ultimamente responsabile del successo dell'intera attività.
- C (Consulted): Le persone che forniscono input o feedback sull'attività.
- I (Informed): Le persone che devono essere tenute al corrente dell'attività, ma non sono direttamente coinvolte nell'esecuzione.

In questo modo, la matrice RACI fornisce una panoramica chiara e concisa dei diversi ruoli e delle relative responsabilità all'interno del progetto.

REGIONE LAZIO

3.3. Analisi

3.3.1. Analisi della qualità

Il processo di analisi cui sottoporre i dati posseduti da un'Amministrazione, con il fine di analizzarne ed incrementarne il livello di qualità, è un'operazione complessa. Per determinare la bontà dei dati è necessario definire delle metriche attraverso le quali misurare la qualità dei dati, processo che, tuttavia, trova la sua complessità nell'identificazione di metriche universalmente valide, in quanto la correttezza dei dati è profondamente legata ai singoli contesti operativi. Lo standard ISO/IEC 25012:2008, divenuto norma italiana UNI ISO/IEC 25012:2014, definisce un insieme di dimensioni specifiche per la caratterizzazione della qualità dei dati:

- **accuratezza** (sintattica e semantica): il dato, e i suoi attributi, rappresenta correttamente il valore reale del concetto o evento cui si riferisce;
- **coerenza**: il dato, e i suoi attributi, non presenta contraddittorietà rispetto ad altri dati del contesto d'uso dell'amministrazione titolare;
- **completezza**: il dato risulta esaustivo per tutti i suoi valori attesi e rispetto alle entità relative (fonti) che concorrono alla definizione del procedimento;
- **attualità** (o tempestività di aggiornamento): il dato, e i suoi attributi, è del "giusto tempo" (è aggiornato) rispetto al procedimento cui si riferisce.

L'ISO/IEC 25012 è stato in seguito integrato mediante lo standard ISO/IEC 25024 ("*Measurement of Data Quality*"), specificando per ciascuna dimensione, la qualità delle opportune metriche (vedi paragrafo con metriche di esempio), in base alle quali è possibile valutare la bontà complessiva dei dati e stimare la loro qualità per tutte le caratteristiche previste dall'ISO/IEC 25012.

R	A	C	I
Referenti operativi delle Strutture regionali	Commissione Tecnica	-	Commissione Tematica

Copia

REGIONE LAZIO

3.3.2. Bonifica

Spesso i dati all'interno dei sistemi informativi o degli archivi di un'amministrazione sono "sporchi" e non rispondenti alle principali raccomandazioni sulla qualità dei dati e la loro standardizzazione (vedi paragrafo in appendice), motivo per cui spesso è necessario effettuare un'attività di bonifica. I processi di bonifica prevedono che il dataset sia corretto mediante confronto con altri dataset correlati e/o analoghi (pubblicati sia dall'ente stesso che da altri), anche mediante attività economicamente onerose, come contattare direttamente i soggetti preposti alla gestione della base dati che presenta errori per correggerli insieme a loro ed effettuare un confronto incrociato (*matching*) con altri dataset. Tali processi hanno il vantaggio di poter pervenire al risultato, ma lo svantaggio di non risolvere il problema alla radice. Infatti, in un arco temporale medio-lungo, i dataset potrebbero nuovamente presentare problemi di qualità: ad esempio, per evoluzione dei contesti (normativi, tecnologici, ecc.), o per errori o modifiche ai dati e al loro processamento.

È essenziale valutare con attenzione le strategie di gestione e manutenzione dei dati. Tale valutazione deve essere orientata non solo alla risoluzione dei problemi esistenti, ma anche alla prevenzione di quelli futuri.

Ecco alcuni passaggi critici da considerare:

- **Analisi delle Cause**: Individuare le cause ricorrenti che generano dati di bassa qualità, inclusi i processi di aggiornamento non efficaci o le lacune nelle normative vigenti.
- **Standardizzazione dei Processi**: Definire standard chiari e precisi per la gestione dei dati, compresi i protocolli di aggiornamento e le procedure di controllo qualità, con l'obiettivo di rendere ogni passaggio del processo il più invariante possibile rispetto a cambiamenti normativi o tecnologici.

Copia

REGIONE LAZIO

- **Integrazione di Controlli Automatici**: Implementare sistemi informativi dotati di meccanismi automatici di validazione dei dati in tempo reale che possano intercettare e segnalare anomalie non appena si verificano.
- **Revisione Periodica e Audit**: Pianificare revisioni periodiche dei dati e audit indipendenti per valutare l'efficacia delle politiche di qualità dei dati e l'aderenza agli standard definiti.
- **Aggiornamento delle Competenze**: Assicurare che il personale sia costantemente aggiornato rispetto alle evoluzioni normative e tecnologiche e che sia formato per gestire adeguatamente i dati secondo le best practice del settore.
- **Retroazione e Miglioramento Continuo**: Sviluppare un meccanismo di feedback che coinvolga gli utenti dei dati, permettendo loro di segnalare problemi e suggerire miglioramenti, e che garantisca l'elaborazione di tali riscontri per il continuo miglioramento dei processi.

R	A	C	I
Referenti operativi delle Strutture regionali	Commissione Tecnica	-	Commissione Tematica

3.3.3. Analisi di processo

Un altro metodo di bonifica è quello basato sui processi, che prevede di analizzare le cause che hanno portato ad errori sui dati e di rivedere i processi di produzione del dato stesso per garantirne la qualità nel tempo.

È preferibile che l'aggiornamento dei processi sia gestito dai Referenti operativi delle Strutture regionali, possibilmente in coordinamento con la commissione tecnica.

REGIONE LAZIO

R	A	C	I
Referenti operativi delle Strutture regionali	Referente Operativo Open Data	Commissione Tecnica	Commissione Tematica

Copia

REGIONE LAZIO

3.4. Arricchimento

I dati possono essere arricchiti attraverso l'integrazione con altri dati e il linking esterno. L'arricchimento è definito dal documento "*Data quality guidelines*[25]" del Publications Office come il concetto *di "collegare i dati da fonti esterne ai set di dati esistenti"* ed è un processo costituito sia dalla standardizzazione che dall'arricchimento vero e proprio.

L'arricchimento dei dati può essere ottenuto anche integrando informazioni da sorgenti esterne - permettendo non solo di correggere e aggiornare i valori esistenti, ma anche di aggiungere nuovi campi informativi e, in alcuni casi, di ridefinire lo schema stesso dei dati per una maggiore precisione e completezza - rendendo in questo modo i dati di origine più significativi e fruibili.

Un'ulteriore attività è quella che si basa sui LOD (*Linked Open Data*), cioè quell'insieme di principi e metodi per pubblicare dati aperti strutturati usando tecnologie web standard, permettendo di collegare e interrogare dati provenienti da diverse fonti, tipicamente fra amministrazioni o istituzioni diverse, ma anche, al limite, all'interno di una stessa amministrazione. Il collegamento o linking dei dati può in effetti aumentarne il valore creando nuove relazioni e consentendo così nuovi tipi di analisi. Tale collegamento è possibile soprattutto attraverso l'uso coerente di identificatori univoci, gli URI e di diversi standard e tecniche, tra cui il framework RDF, e risorse come i vocabolari controllati.

I vocabolari controllati, definiti[26] come "un modo comune e condiviso per organizzare codici e nomenclature ricorrenti in maniera standardizzata e normalizzata", assegnano ad ogni concetto un identificatore univoco e persistente (URI), in modo che quel concetto venga referenziato in maniera non ambigua, e garantiscono, inoltre, la gestione coerente delle diverse versioni. Possono essere associate oltre alle etichette,

[25] https://op.europa.eu/en/publication-detail/-/publication/b601d9cc-b3c0-11ec-9d96-01aa75ed71a1
[26] Articolo 4 del Piano Triennale per l'informatica nella Pubblica Amministrazione 2017-2019

REGIONE LAZIO

definizioni e descrizioni anche in diverse lingue. In caso di dati di elevato valore, per specifiche categorie tematiche (ovvero osservazione della terra e ambiente, statistica, imprese e proprietà delle imprese, mobilità) i set di dati devono utilizzare, ove possibile, vocabolari controllati e tassonomie documentati pubblicamente e riconosciuti nell'Unione o a livello internazionale.

Ad esempio

1. **Osservazione della Terra e Ambiente**:
    - INSPIRE (*Infrastructure for Spatial Information in the European Community*): Un'iniziativa dell'Unione Europea per creare una infrastruttura europea per i dati spaziali ambientali per supportare le politiche ambientali.
    - OGC (*Open Geospatial Consortium*) Standards: Standard internazionali per l'interoperabilità dei servizi e dei dati geospaziali.
2. **Statistica**:
    - SDMX (*Statistical Data and Metadata eXchange*): Uno standard internazionale per lo scambio di dati statistici e metadati tra istituzioni.
3. **Mobilità**:
    - GTFS (*General Transit Feed Specification*): Un formato comune per la condivisione di dati sui trasporti pubblici, che permette di descrivere orari, percorsi e altre informazioni sui servizi di trasporto.

R	A	C	I
Referenti operativi delle Strutture regionali	Commissione Tecnica	-	Commissione Tematica

Copia

REGIONE LAZIO

3.5. Modellazione e Documentazione

La preparazione dei dati inizia con la definizione di come sono strutturati (sintassi) e cosa significano (semantica). Questo passaggio rende più facile usare i dati, migliora la loro qualità, e ne aumenta il valore, poiché si riduce il rischio di interpretarli in modo errato fornendo un contesto chiaro.

Se i dati non esistono ancora, questo step diventa il primo del processo. Si inizia creando un modello che spiega come i dati saranno organizzati e cosa rappresenteranno, definendo entità, attributi e le loro relazioni: è la fase di modellazione.

Al fine di assicurare una corretta interpretazione dei dati da parte degli utenti e dei sistemi automatici, è opportuno condividere gli schemi di tali dati, che possono essere specificati separatamente e forniscono una descrizione della propria struttura e sintassi. Ci sono delle modalità standard di farlo, che sono quelle riportate nel capitolo "Modellazione e documentazione" delle Linee Guida Open Data nazionali.

Il "Piano Triennale ICT 2017-2019[27]" descrive un modello dati condiviso o un'ontologia come una rappresentazione dettagliata e precisa di un particolare settore. Per assicurare coerenza nei documenti, si sottolinea l'importanza dell'uso di modelli condivisi nella comunicazione tra diverse entità. Questo approccio aiuta a standardizzare e uniformare le informazioni, rendendo più facile l'interoperabilità tra soggetti diversi.

Le "Linee Guida sull'interoperabilità tecnica[28]" hanno già stabilito alcuni criteri per questi modelli dati. Come indicato in questi documenti, per definire le entità da condividere, i soggetti devono:

[27] https://www.agid.gov.it/it/agenzia/strategia-quadro-normativo/piano-triennale
[28] https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_interoperabilit_tecnica_pa.pdf

REGIONE LAZIO

- Identificare i settori di interesse e determinare le entità da rappresentare, insieme alle loro caratteristiche specifiche;
- Controllare se le entità per ogni settore sono già presenti nella rete nazionale di ontologie e vocabolari, che si possono trovare nel Catalogo Nazionale per la semantica dei dati.

In aggiunta, per i dati territoriali, i modelli dati da usare sono quelli definiti nella Direttiva INSPIRE[29] e le estensioni nazionali della stessa.

R	A	C	I
Referenti operativi delle Strutture regionali	Commissione Tecnica	-	Commissione Tematica

Copia

[29] https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A32007L0002&qid=1658755217938

REGIONE LAZIO

3.6. Validazione

Lo scopo della validazione è quello di assicurare un certo livello di qualità ai dati[30]. Questa fase deve essere effettuata prima della pubblicazione e dell'utilizzo dei dati stessi. La validazione può essere definita "*come un processo che assicura la corrispondenza dei dati finali (pubblicati) con una serie di caratteristiche qualitative*"[31].

La fase di validazione può essere un ulteriore passaggio per la verifica dei dati a valle delle altre operazioni effettuate prima della pubblicazione, per esempio per l'arricchimento. Può essere anche intesa come la prima e unica fase del processo per la verifica della qualità nel caso di nuovi dati che, quindi, non rientrerebbero nella fase di ricognizione e nella conseguente analisi.

In ogni caso il Responsabile Operativo Open Data, prima della pubblicazione definitiva di un dataset deve assicurarsi che esso rispetti al massimo possibile i punti chiave indicati di seguito:

- **Accuratezza dei dati**: Verificare la correttezza e l'accuratezza dei dati per garantire che siano affidabili.
- **Completezza**: Assicurarsi che il dataset sia il più completo possibile, includendo tutti i dati rilevanti.
- **Attualità**: Controllare che i dati siano aggiornati al momento della pubblicazione.
- **Consistenza**: Verificare la coerenza interna dei dati, assicurandosi che non ci siano contraddizioni o incongruenze.

[30] Di analisi di qualità dei dati si è già accennato nella fase post ricognizione dei dati.
[31] Definizioni riportate nel documento "Methodology for data validation 2" (https://cros-legacy.ec.europa.eu/system/files/ess_handbook_-_methodology_for_data_validation_v2.0_-_rev2018_0.pdf)

Copia

REGIONE LAZIO

- **Formattazione standardizzata**: Controllare che i dati siano formati in modo standard per facilitare l'uso e l'interpretazione.
- **Metadata adeguati**: Fornire metadati completi e accurati che descrivano il dataset, inclusi la fonte, la data di raccolta, e la metodologia utilizzata.
- **Accessibilità**: Assicurarsi che il dataset sia facilmente accessibile e utilizzabile dal pubblico.
- **Interoperabilità**: Verificare che il dataset sia compatibile con altri set di dati e sistemi, favorendo l'integrazione.
- **Rispetto della privacy**: Garantire che il dataset rispetti le normative sulla privacy e sulla protezione dei dati.
- **Licenze adeguate**: Assicurarsi che il dataset sia accompagnato da una licenza chiara che ne specifichi le condizioni di uso.
- **Non ambiguità**: Verificare che i dati siano presentati in modo chiaro e non ambiguo.
- **Rispettare gli standard internazionali**: Assicurarsi che il dataset aderisca agli standard internazionali rilevanti, se applicabili.

Per approfondire il tema della validazione e di possibili metodi si rimanda alla pubblicazione "*Methodology for data validation 2.0*[32]", citata nella Linee Guida Open Data nazionali, la quale è stata sviluppata nell'ambito dello *European Statistical System (ESS)* e si concentra su metodologie e pratiche per la validazione dei dati, fornendo metodologie standardizzate e approcci volti ad assicurare l'accuratezza e l'affidabilità dei dati.

R	A	C	I
Referente Operativo Open Data	Referente Operativo Open Data	Commissione Tecnica	Commissione Tematica

[32] https://cros-legacy.ec.europa.eu/system/files/ess_handbook_-_methodology_for_data_validation_v2.0_-_rev2018_0.pdf

Copia

REGIONE LAZIO

3.7. Pubblicazione

Prima di procedere alla pubblicazione è necessario avere raccolto tutte le informazioni utili per definire i metadati, definire le politiche di accesso, il modello di licenza da applicare e identificare i canali per la pubblicazione, anche in base alla tipologia dei dati. Tra i possibili canali distinguiamo tra:

- accesso diretto (del singolo dataset e/o in blocco)
- portale dati
- API

R	A	C	I
Referenti operativi delle Strutture regionali	Referenti operativi delle Strutture regionali	-	Commissione Tecnica Commissione Tematica

3.7.1. Metadatazione e uso delle API

Come detto precedentemente, la metadatazione è cruciale: i metadati certificano le caratteristiche del dato[33]. Per i metadati descrittivi generali, ovvero non dipendenti dalle tipologie di dati, deve essere applicato il profilo nazionale DCAT-AP_IT, rispettando le obbligatorietà, le raccomandazioni e seguendo gli esempi così come definiti nella relativa specifica e ontologia.

[33] "Linee Guida recanti regole tecniche per l'attuazione del decreto legislativo 24 gennaio 2006, n. 36 e s.m.i. relativo all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico", cap.4, par. 4.6.

REGIONE LAZIO

I dati dinamici e le serie di dati di elevato valore devono essere resi disponibili attraverso delle API[34] basate sui seguenti principi: disponibilità, stabilità, manutenzione per tutto il ciclo di vita, uniformità di utilizzo e delle norme, facilità d'uso e sicurezza[35].

3.7.2. Aspetti architetturali relativi alla pubblicazione dei dataset aperti

Per quanto riguarda gli aspetti architetturali si distinguono due livelli:

- livello di *front-end*;
- livello infrastrutturale.

Il livello di *front-end* svolge una funzione di presentazione e può essere sia un sito Web, sia una sezione in un sito esistente. In questa parte rientrano tutti quegli strumenti che consentono di dare massima visibilità ai dati disponibili e di interagire in maniera semplice e intuitiva con gli utenti stessi, per esempio per capire quali dati sono di loro interesse, quali nuovi dati sono richiesti, quali suggerimenti vogliono dare per migliorare la qualità dei dati. Il livello di presentazione si completa con l'interfaccia di accesso via web per interrogazioni puntuali sui dati e metadati. Questa ha come obiettivo quello di aumentare l'interazione machine-to-machine attraverso il dispiegamento di una piattaforma di esposizione dati basata su API di servizio.

In generale, si raccomanda di:

- assegnare ai dataset nomi autoesplicativi per comprenderne il contenuto;
- fornire, ove possibile, descrizioni testuali dei dataset;
- mettere in evidenza la licenza in uso in forma "*human and machine-readable*";

[34] Il catalogo delle API, in applicazione dell'art. 50 ter del CAD, è reso disponibile attraverso la Piattaforma Digitale Nazionale Dati a cui si applicano le indicazioni delle "*Linee Guida sull'infrastruttura tecnologica della Piattaforma Digitale Nazionale Dati per l'interoperabilità dei sistemi informativi e delle basi di dati*" adottate con la Determinazione di AgID n. 627/2021.

[35] In tema di API, con la Determinazione di AgID n. 547/2021 sono state adottate le "Linee Guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni" e le "Linee Guida Tecnologie e standard per la sicurezza dell'interoperabilità tramite API dei sistemi informatici", che definiscono le modalità con cui le Pubbliche Amministrazioni implementano le proprie API quale elemento tecnologico di base del modello di interoperabilità.

REGIONE LAZIO

- fornire, ove possibile, strumenti di visualizzazione e navigazione, anche georiferita, dei dati, che possano facilitare la lettura degli stessi;
- fornire, ove possibile, statistiche di uso, accesso e produzione;
- fornire notifiche di cambiamenti nel sito web, di aggiornamenti ai dataset;
- fornire strumenti per rendere agevoli le interrogazioni, anche per non esperti [36].

Il livello infrastrutturale è rappresentato dall'infrastruttura che ospita i dati e i metadati. Nel caso di dati aperti, tenuto conto della loro natura intrinseca, ovvero dati tipicamente non riferibili a singole persone (e per i quali solitamente non si richiede il soddisfacimento di specifici requisiti di protezione dei dati personali), tecnologie basate sul paradigma del cloud computing pubblico [37] possono essere facilmente impiegabili al fine di ospitare le infrastrutture per la pubblicazione di dati aperti. AgID ha adottato, con la Determinazione n. 628/2021, il Regolamento che definisce i requisiti minimi per le infrastrutture digitali e le caratteristiche e le modalità di qualificazione e migrazione dei servizi cloud.

Copia

[36] Nel caso dei dati dei livelli 4 e 5 non si può pubblicare solo dataset RDF ma è bene mettere in evidenza la presenza dello SPARQL endpoint (cioè, un servizio Web che accetta interrogazioni SPARQL, le risolve e restituisce i risultati in output), pubblicando il link di accesso, fornendo altresì un ampio insieme di "query" di esempio che con pochi click possono essere eseguite producendo risultati disponibili in diversi formati di più comune utilizzo soprattutto nell'ambito delle comunità open data (per es., CSV, JSON, XML).

[37] In tema di cloud, la Strategia Cloud Italia definita dal Dipartimento per la trasformazione digitale e dall'Agenzia per la cybersicurezza nazionale (ACN), contiene gli indirizzi strategici per il percorso di migrazione verso il cloud di dati e servizi digitali della Pubblica Amministrazione.

REGIONE LAZIO

3.7.3. Uso di identificatori univoci e persistenti

Nei requisiti per i dati della ricerca volti a rendere tali dati conformi ai principi FAIR si è fatto riferimento più volte a identificatori univoci e persistenti. Tali identificatori sono generalmente rappresentati dagli URI (*Uniform Resource Identifier*), una sequenza di caratteri che identifica una risorsa astratta o fisica. Gli URI [38] devono essere persistenti e dereferenziabili. Nella gestione degli URI è opportuno utilizzare il meccanismo di "*content negotiation*" che consente di rendere disponibile, allo stesso URI, diverse rappresentazioni di una risorsa in caso di molteplici rappresentazioni possibili (ad esempio, URI che rappresentano sia entità del web semantico sia risorse web). Così come è una buona prassi utilizzare sempre URI "dereferenziabili", come già indicato prima, cioè URI che restituiscono una rappresentazione web (es. una pagina web informativa) di una risorsa.

Una indicazione utile per strutturare adeguatamente un URI è la seguente:

- strutturare l'URI come segue:

  http://{dominio}/{tipo}/{concetto}/{riferimento}

- includere nell' URI i seguenti elementi:
  - dominio: il dominio Web su cui reperire la risorsa
  - tipo: l'elemento che specifica il tipo di risorsa. Dovrebbe poter assumere un numero limitato di valori come "doc" se la risorsa identificata è un documento descrittivo, "set" se la risorsa è un dataset, "id" o "item" se la risorsa è un oggetto del mondo reale

[38] Per la creazione di URI persistenti sono da evitare quelli che contengano:

- nome del progetto/ufficio/unità amministrativa che detiene la risorsa per evitare problemi derivanti dalla fine del progetto stesso o fusioni o chiusure di uffici nell'organizzazione;
- numeri di versione;
- identificatori esistenti che in passato sono stati utilizzati per identificare risorse differenti;
- riferimenti generati in modo automatico e incrementale a meno che non vi sia la garanzia che il processo non venga mai più ripetuto o, se ripetuto, generi sicuramente gli stessi identificatori per gli stessi dati di input;
- stringhe rappresentanti "query" a database;
- estensione del file;

Copia

- concetto: il tipo di un oggetto del mondo reale
- riferimento: lo specifico elemento, termine o concetto che rappresenta la risorsa

3.7.4. Politiche di accesso

Altro aspetto importante da considerare sono eventuali forme di aggregazione dei dati e restrizioni di accesso, che hanno anche un impatto sulla scelta della licenza, per cui si rimanda al capitolo 4 della presente Guida.

Sebbene sia sconsigliato restringere l'accesso ai dati o procedere con la pubblicazione di aggregazioni degli stessi (in generale non è opportuno che l'esposizione del dato lavorato avvenga senza che sia stato pubblicato prioritariamente il dato grezzo), esistono casi in cui i dati possono essere diffusi solo in forma anonima (ad esempio i redditi), ossia a un livello di aggregazione tale da impedire di identificare le persone cui i dati si riferiscono.

Copia

REGIONE LAZIO

3.8. Coinvolgimento dell'utenza finale

Il coinvolgimento dell'utenza finale rappresenta un aspetto importante del processo di apertura dei dati pubblici. La Regione Lazio, dunque, si pone l'obiettivo di coinvolgere le persone, le aziende, gli enti di ricerca, le professioniste e i professionisti che utilizzeranno i dati pubblicati nel portale regionale.

Tale procedura è importante perché:

- Migliora la qualità dei dati: soggetti esterni possono segnalare errori, dare suggerimenti per migliorare dati e metadati.
- Aumenta il valore dei dati: attraverso nuove analisi, visualizzazioni, applicazioni realizzate da sviluppatori e aziende private.
- Stimola l'innovazione: i dati aperti abilitano lo sviluppo di servizi digitali utili a cittadini e imprese.
- Promuove la trasparenza: i cittadini possono monitorare più da vicino le attività della PA e fornire riscontri.
- Coinvolge la comunità: i dati diventano un bene comune da tutelare e far crescere insieme.
- Raccoglie nuovi fabbisogni: il confronto con l'esterno serve a capire come migliorare nel rilascio dei dati più utili.
- Migliora l'efficienza: i suggerimenti del pubblico possono portare a risparmi e semplificazioni.

Per questo motivo sarò necessario:

- che gli indirizzi di posta elettronica di contatto, inseriti nei dataset, siano presidiati da chi li detiene;
- documentare il funzionamento delle API, anche con pochi esempi di "primo utilizzo":

Copia

REGIONE LAZIO

- dare evidenza a tutti gli usi dei dati fatti da studenti e studentesse, aziende, dal mondo della ricerca e del giornalismo, ecc., e mettere a disposizione una modalità per segnalarli al Referenti operativo della Struttura regionale;
- prevedere un incontro annuale con l'utenza finale per discutere le esigenze degli utenti e fornire informazioni sulle attività di apertura dei dati pubblici della Regione Lazio.

R	A	C	I
Commissione Tematica	Commissione Tematica	Referente Operativo Open Data	Responsabile Strategia Open Data

Copia

REGIONE LAZIO

4. Buone pratiche

In questo paragrafo sono elencate delle buone pratiche propedeutiche alla pubblicazione di dati online, alide sia per chi deve predisporre dei dati per la pubblicazione sul portale open data regionale, che per chi gestirà la pubblicazione dei dati. Si farà riferimento al formato CSV, perché molto diffuso, ma la gran parte delle pratiche hanno valore generale.

4.1. Definizione "Note" del tracciato record

Ad ogni risorsa del portale è associato un file con il tracciato record dei campi, in cui ogni attributo è associato al relativo tipo di campo. È essenziale che sia valorizzato anche il campo "Note" del tracciato del *record*, perché in molti casi il nome del campo da solo non restituisce le informazioni necessarie a una sua corretta interpretazione.

Se i valori di un campo sono quelli di liste di codici definiti (ad esempio i codici Istat dei Comuni), inserire il riferimento che consenta di risalire alla definizione e al significato di questi.

4.2. Codifica dei caratteri

La codifica dei caratteri – *encoding* – è il punto di ingresso per la lettura dei dati ed evita di incorrere in problemi di visualizzazione o interpretazione dei dati, specialmente quando si utilizzano caratteri speciali o alfabeti non latini.

È bene scegliere sempre l'UTF-8, in quanto è quella più "standard" per la gran parte delle applicazioni e dei linguaggi di *scripting* per l'analisi dei dati.

In ogni caso, è sempre da documentare, in quanto esistono formati (come il CSV) in cui l'informazione di *encoding* non è inserita all'interno del file.

4.3. Valori sconosciuti o nulli

Tutte le celle di una tabella dovrebbero essere valorizzate e dovrebbe essere scelto un codice comune per i dati sconosciuti. Diamo come scelta predefinita che la cella

Copia

vuota rappresenti un *NULL value*, un valore "sconosciuto" o "non disponibile"; se si usano allo scopo valori come "NA", "-" o "N/A", specificarlo nei metadati della risorsa.

4.4. Scelta dei nomi dei campi

I nomi dei campi o delle colonne in una tabella di dati dovrebbero essere comprensibili dalle persone. In molti casi, le colonne delle tabelle di dati mantengono i nomi assegnati dai sistemi di gestione delle banche dati, di solito soggetti a convenzioni tecniche che sono difficili da capire per le persone.

Si raccomanda di:

- usare nomi brevi (non più di 20 caratteri), ma tenendo sempre presente che il risparmio di caratteri non deve portare a un'errata interpretazione del nome del campo;
- evitare di usare abbreviazioni;
- usare solo caratteri ASCII minuscoli (a-z; 0-9);
- non usare caratteri speciali (per esempio äüöàéèê, ecc.);
- non includere accenti o segni di punteggiatura;
- usare i trattini bassi _ per separare le parole che compongono i nomi delle colonne invece degli spazi bianchi;
- evitare l'uso di codici, e se assolutamente necessario, dovrebbero essere illustrati nel dizionario dei dati che documenta il set di dati.

4.5. Rappresentazione Wide e Long

Quando si progettano strutture di dati tabulari è consigliabile evitare la crescita orizzontale dei valori, e disporre i valori di una variabile per colonna e non per riga; in letteratura sono i cosiddetti *Tidy Data*[39].

[39] Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10), 1–23. https://doi.org/10.18637/jss.v059.i10

REGIONE LAZIO

Qui sotto un esempio in cui il tipo di guasto di un modello di automobile è riportato in orizzontale (si dice "*wide*", ovvero "largo"). Per ogni tipo una colonna.

modello_auto	*guasto_radiatore*	*guasto_carburatore*	*guasto_sospensione*	*guasto_frizione*
chevrolet chevelle malibu	0	7	1	0
buick skylark 320	1	2	2	2

Mentre è preferibile e consigliato, avere una riga per ogni osservazione, e una colonna per ogni variabile. Quindi l'esempio si trasforma nella tabella seguente, in cui è stato aggiunto il campo con il tipo di guasto.

modello_auto	*tipo_guasto*	*quantita_guasti*
chevrolet chevelle malibu	Radiatore	0
chevrolet chevelle malibu	Carburatore	7
chevrolet chevelle malibu	sospensione	1
chevrolet chevelle malibu	Frizione	0
buick skylark 320	Radiatore	1
buick skylark 320	Carburatore	2
buick skylark 320	sospensione	2
buick skylark 320	Frizione	2
plymouth satellite	Radiatore	0
plymouth satellite	Carburatore	4
plymouth satellite	sospensione	4
plymouth satellite	Frizione	1

4.6. Subtotali, totali o raggruppamenti

Nessuna riga o colonna con totali o subtotali dovrebbe essere inclusa, a meno che non sia assolutamente necessario, mantenendo il massimo livello possibile di disaggregazione dei dati.

4.7. Tipi di dati

I valori in una tabella di dati devono essere formattati secondo il tipo di dati relativo. In particolare, i numeri dovrebbero essere sempre nei formati/tipi di cella "Numero", i campi di tipo testo dovrebbero essere nel formato/tipo di cella "Testo" e i campi di tipo data dovrebbero essere nel formati/tipi di cella "Data".

Questo ha un impatto anche in formati come il CSV, che non hanno la definizione del tipo di campo. Ad esempio, il valore numerico "12256,12" non deve essere

REGIONE LAZIO

rappresentato in un CSV con il separatore delle migliaia - "12.256,12"- perché il tipo di campo così passerebbe da numerico a testuale.

A seguire alcune raccomandazioni:

- Nei campi di testo, non devono esserci spazi bianchi a inizio e fine cella, o doppi spazi nelle altre parti;
- Nei campi numerici non deve essere presente il separatore delle migliaia;
- Non inserire l'unità di misura nei campi numerici, ma soltanto il valore;
- Nei campi numerici inserire il numero dei decimali coerenti con le modalità di acquisizione dei valori;
- Per campi con date, o con data e ora, è molto consigliabile usare la rappresentazione dello standard "ISO 8601[40]" (ad esempio AAAA-MM-GG, ovvero 2023-06-28). In questo modo i valori saranno letti correttamente da qualsiasi applicazione e/o linguaggio di programmazione;
- Per i campi con coppie di coordinate geografiche, per favorire l'usabilità immediata, è consigliabile utilizzare la rappresentazione in gradi decimali con sistema di riferimento "EPSG:4326[41]".

4.8. Standardizzazione dei valori dei campi

L'utilizzo di classificazioni, vocabolari e più in generale di metadati "standard" permette di combinare tra loro dati provenienti da fonti differenti, abilitandone in questo modo l'interoperabilità.

Vanno considerati tecnicamente corretti quei dati che:

- usano la stessa codifica e "normalizzazione" per lo stesso tipo di dati, pubblicati in diversi dataset di un catalogo, per esempio, gli indirizzi sono sempre pubblicati con la stessa struttura, tipo, formato;

[40] Standard internazionale per la rappresentazione di date e orari
[41] WGS84 - World Geodetic System 1984 https://epsg.io/4326

Copia

REGIONE LAZIO

- in cui la codifica si basa su alcuni standard comuni proposti e utilizzati da altre organizzazioni nazionali o internazionali come, ad esempio, Schema [42], EUROSTAT o ISTAT.

In generale, per i campi dove è possibile farlo, è necessario aggiungere anche la rappresentazione in termini di codice del valore e non soltanto di etichetta. Un esempio è quello dei nomi dei Comuni: è essenziale aggiungere anche il codice Istat (sia "Latina", che "059011").

La normalizzazione è necessaria per evitare ad esempio di pubblicare dataset distinti con tipi di campo omologhi (ad esempio la temperatura), ma i cui valori sono espressi con unità di misura e/o precisione differente.

Per questa standardizzazione è utile fare riferimento a "panieri di dati" esistenti (vedi capitolo 5), o definire all'occorrenza degli schemi dati da condividere con gli uffici regionali.

Copia

[42] Il catalogo nazionale per lo scambio di dati e informazioni tra pubbliche amministrazioni https://schema.gov.it/

REGIONE LAZIO

5.Paniere dei dati

Il Paniere dinamico di dataset [43] a cura di AgID rappresenta un elemento di monitoraggio delle politiche di open data nazionali: viene utilizzato per la definizione dello stato di avanzamento del paese come rappresentato nell'*Open Data Maturity* in Europe e nell'indice DESI[44], sempre con riferimento al tema dell'Open data.

È usato come riferimento per la scelta e individuazione dei dataset da pubblicare. La sua ultima versione a cura di AgID è quella del 2021; al suo interno contiene il foglio "Paniere_base_regionale_2021", che è quello da tenere come riferimento. In questo elenco sono presenti per ogni dataset, la categoria DCAT-AP_IT, il nome e la descrizione.

Nel solco del suddetto paniere si sono sviluppate delle iniziative interessanti di Regione Lombardia e Regione Piemonte, che hanno costruito un paniere regionale basato sullo studio di quanto pubblicato da altre Regioni e sul confronto con altri enti locali (Comuni, Città metropolitane, ecc.). In questo paniere, rispetto a quello di AgID, è stata aggiunta la definizione del "tracciato record", ovvero dello schema dati. Per ogni dataset:

- Nome
- Descrizione
- Tipo
- Obbligatorietà

Avere uno schema dati di riferimento, consente di superare un problema tipico: non potere utilizzare in modo automatico lo stesso dataset pubblicato da due pubbliche amministrazioni distinte. Perché solitamente sarà diverso il numero dei campi, il nome dei campi, la tipologia. Con l'utilizzo di uno schema dati di riferimento, enti locali che

[43] https://www.dati.gov.it/monitoraggio/paniere-dataset
[44] https://digital-strategy.ec.europa.eu/en/policies/desi

Copia

REGIONE LAZIO

generano lo stesso dataset, ad esempio le aree verdi di un Comune, renderanno disponibili dei dati che potranno essere integrati tra loro.

L'elenco dei dataset prodotto da queste regioni è riportato di seguito e può essere usato come ulteriore riferimento.

Tema	Nome Dataset
AMBIENTE	Aree Verdi
AMBIENTE	Aree Verdi Informazioni
AMBIENTE	Carta degli incendi
AMBIENTE	Elenco degli impianti di depurazione
COMMERCIO	Pratiche SUAP
COMMERCIO	Pubblici esercizi
COMMERCIO	Servizi alla persona
CULTURA	Beni artistici e architettonici
CULTURA	Eventi
ENERGIA	Albo provinciale manutentori impianti termici
MOBILITA' E TRASPORTI	Aree di sosta per disabili
MOBILITA' E TRASPORTI	Aree pedonali
MOBILITA' E TRASPORTI	Autoscuole
MOBILITA' E TRASPORTI	Bike sharing
MOBILITA' E TRASPORTI	Orario Trasporto Pubblico
MOBILITA' E TRASPORTI	Parcheggi
MOBILITA' E TRASPORTI	Piste ciclabili
MOBILITA' E TRASPORTI	Punti di ricarica veicoli
MOBILITA' E TRASPORTI	Varchi ZTL
SICUREZZA	Controlli in cantiere effettuati dalla Polizia Locale
SICUREZZA	Incidenti stradali
SICUREZZA	Principali infrazioni al Codice della Strada
SICUREZZA	Provvedimenti adottati dalla Polizia Locale in materia di commercio
STATISTICA	Condizione lavorativa
STATISTICA	Risultati elezioni
STATISTICA	Risultati elezioni comunali
STATISTICA	Risultati elezioni nazionali
STATISTICA	Risultati elezioni regionali
TERRITORIO	Cantieri stradali
TERRITORIO	Centri revisione auto
TERRITORIO	Elenco beni confiscati
TERRITORIO	HotSpot Wifi
TERRITORIO	Pratiche edilizie
TERRITORIO	Rete viaria - Numeri civici

REGIONE LAZIO

Tema	Nome Dataset
TERRITORIO	Rete viaria - Toponimi stradali
TRASPARENZA	Elenco canoni di locazione
TRASPARENZA	Elenco immobili di proprietà
TRASPARENZA	Enti controllati – Enti di diritto privato controllati
TRASPARENZA	Enti controllati – Enti pubblici vigilati
TRASPARENZA	Enti controllati – Società partecipate
TRASPARENZA	Monitoraggio procedimenti
TRASPARENZA	Provvedimenti degli organi di indirizzo politico
TURISMO	Accompagnatori turistici

Per ulteriori riferimenti:

- Regione Lombardia, Standard dei dataset del 'paniere' regionale per gli enti locali;
- Regione Piemonte, Standard dei dataset del 'paniere' regionale per gli enti locali.

Copia

REGIONE LAZIO

6.Licenze

Una volta accertata l'assenza di vincoli di proprietà intellettuale che impediscano l'apertura dei dati, è necessario identificare una licenza che ne definisca le condizioni e le modalità di riutilizzo da parte di chiunque, anche per scopi commerciali, conformemente all'articolo 8 del D.Lgs. 36/2006. Ci sono molte licenze che si prestano al riuso dei dati aperti ma, prima della scelta del tipo specifico di licenza da applicare, è bene considerare che tale scelta comporterà, in sostanza, la classificazione dei dati all'interno di uno di questi gruppi:

- **Dati di Pubblico Dominio**: sgrava l'Ente da ogni responsabilità su eventuali rielaborazioni sbagliate realizzate da terzi;
- **Dati con vincolo di attribuzione**: riconosce all'Ente lo sforzo impiegato nella creazione (o "paternità") dei dati, e richiede a chi lo riutilizza di citare l'Ente nell'opera derivata (vedi definizione di opera nel glossario);
- **Dati con vincolo di condivisione** allo stesso modo: richiede che il risultato del riuso venga condiviso con la stessa licenza, con impiego comune per tutti quei dati che sono stati raccolti in modalità "*crowdsourced*", ovvero con l'aiuto dei cittadini in maniera volontaria (segnalazioni di problemi, idee per lo sviluppo, ecc.).

Tra le varie scelte possibili per definire le condizioni e limitazione per il riutilizzo, le più utilizzate per il rilascio di dati aperti sono:

CC0: consente a chiunque sia proprietario di diritti d'autore o di altri contenuti protetti dalle norme sulle banche dati di rinunciare a tali diritti sulle loro opere e, quindi, lasciarli nel modo più completo possibile al pubblico dominio, in modo che altri possano liberamente sviluppare, migliorare e riutilizzare le opere per altri scopi senza le restrizioni relative al diritto d'autore o al diritto di protezione delle banche dati. Associare un proprio lavoro con questa licenza significa dedicare l'opera al pubblico dominio attraverso la rinuncia a tutti i diritti protetti dal diritto d'autore, inclusi i diritti

Copia

REGIONE LAZIO

ad esso connessi, in tutto il mondo, nei limiti consentiti dalla legge. È possibile quindi copiare, modificare, distribuire ed eseguire l'opera, anche a fini commerciali, il tutto senza chiedere il permesso.

CC BY: consente a chiunque di riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire, modificare o usare l'opera, anche per fini commerciali. L'autore conserva il diritto di rivendicare la paternità dell'opera a suo nome (valgono come nome lo pseudonimo, il nome d'arte, la sigla o il segno convenzionale, che siano notoriamente conosciuti come equivalenti al nome vero) e si richiede all'utilizzatore di attribuire la paternità dell'opera nei modi indicati dall'autore o da chi ha dato l'opera in licenza, in modo tale da non suggerire che essi avallino l'uso che si fa dell'opera.

IODL 2.0: prevede che l'utente possa liberamente consultare, estrarre, scaricare, copiare, pubblicare, distribuire e trasmettere le informazioni, creare un lavoro derivato, per esempio attraverso la combinazione con altre informazioni (mashup), includendole in un prodotto o sviluppando un'applicazione informatica che le utilizzi come base di dati. In cambio, all'utente è chiesto di indicare la fonte delle informazioni e il nome del soggetto che fornisce il dato, includendo, se possibile, un link alla licenza; inoltre (nel caso di uso della versione precedente IODL 1.0) veniva richiesto all'utente di pubblicare e condividere gli eventuali lavori derivati con la stessa licenza o con altra licenza aperta, ritenuta compatibile.

CC BY-SA: oltre alle condizioni di CC BY, si aggiunge che l'opera modificata deve essere rilasciata secondo le stesse condizioni scelte dall'autore originale. Senza pregiudizio dei diritti esistenti sull'opera originaria, sono altresì protette le elaborazioni di carattere creativo dell'opera stessa, quali le traduzioni in altra lingua, le trasformazioni da una in altra forma letteraria od artistica, le modificazioni ed aggiunte che costituiscono un rifacimento sostanziale dell'opera originaria, gli adattamenti, le riduzioni, i compendi o le variazioni non costituenti opera originale.

Copia

REGIONE LAZIO

Esistono altre licenze promosse dall'Open Data Commons[45] (un progetto dell'Open Knowledge Foundation[46]) create in modo specifico per l'uso con i dati aperti, meno diffuse in Italia: ODC-BY, ODbL, ODC-PDDL. Fatti salvi i limiti di legge inderogabili, le licenze da adottare vanno individuate dal responsabile Open Data in modo tale che l'Ente licenziante non si assuma alcuna responsabilità per qualsiasi costo, onere, spesa o danno di qualsiasi natura che dovessero derivare dall'esercizio delle facoltà concesse in licenza, compreso l'utilizzo o il mancato utilizzo del documento da parte del licenziatario o da parte di terzi. A ciascun dataset verrà abbinata la licenza con cui viene pubblicato e questa dovrà successivamente essere accessibile in lingua italiana. Il Responsabile Open Data deve prediligere una licenza leggibile e riconoscibile non solo agli utenti nazionali, ma anche agli utenti internazionali e che garantisca la massima possibilità di riutilizzo dei dati, a meno di specificità che richiedano l'adozione di licenze meno aperte (ad es.: uso dei dati a pagamento).

Copia

[45] https://opendatacommons.org/
[46] https://okfn.org/

REGIONE LAZIO

7.Conclusioni

Si vogliono sottolineare alcuni aspetti chiave di queste Linee Guida Regionali Open Data:

- **Complementarità con la Normativa Nazionale**: questo documento non sostituisce le direttive nazionali sull'Open Data, ma mira a integrarle fornendo indicazioni specifiche adatte al contesto regionale.
- **Riferimento alle Fonti Ufficiali**: si invita gli utenti a consultare sempre le disposizioni e le Linee Guida Open Data nazionali. Questo assicura coerenza e allineamento con le norme vigenti a livello nazionale e internazionale.
- **Supporto all'Amministrazione Regionale**: le presenti Linee Guida sono state elaborate per assistere l'amministrazione regionale nell'implementazione di politiche di Open Data efficaci, promuovendo un processo di apertura dei dati chiaro e trasparente.
- **Aggiornamento**: considerando la natura sempre in evoluzione del contesto normativo e applicativo dell'Open Data, questo documento è pensato per essere regolarmente aggiornato. Sarà rivisto e adattato in risposta a modifiche normative o a nuove esigenze applicative.

Queste Linee Guida sono un supporto progettato per guidare e facilitare l'attuazione delle politiche di Open Data nella regione.

Copia

REGIONE LAZIO

8. Appendice

8.1. Metadati: principali obbligatorietà e raccomandazioni

Nelle Linee guida per i cataloghi dati di AgID[47] sono definiti i dettagli su tutti gli elementi obbligatori e raccomandati per la definizione di un catalogo dei dati aperti.

Qui a seguire in evidenza quelli più "quotidiani" e obbligatori per la creazione e manutenzione dello stesso.

- Per il **Catalogo**
  - Titolo
  - Descrizione
  - Editore
  - Data ultima modifica
  - Dataset (ogni catalogo deve contenere almeno un dataset)
- Per un **Dataset**
  - Identificativo del dataset
  - Titolo del dataset
  - Descrizione del dataset
  - Ultima modifica del dataset
  - Temi del dataset
  - Titolare del dataset
  - Frequenza aggiornamento dataset
  - Distribuzione del dataset (ogni dataset deve contenere almeno una distribuzione/risorsa)
- Per una **Distribuzione** (risorsa)
  - Formato Distribuzione
  - URL di accesso

[47] https://docs.italia.it/italia/daf/linee-guida-cataloghi-dati-dcat-ap-it/it/stabile/index.html

Copia

REGIONE LAZIO

- Licenza

Copia

REGIONE LAZIO

8.2. Misure per la valutazione qualità Standard ISO/IEC 25024

La **raccomandazione 8** delle Linee Guida Regionali Open Data prevede che sia garantita, per tutti i dati in generale e per quelli resi disponibili per il riutilizzo, in particolare, il rispetto almeno delle **quattro caratteristiche di qualità dei dati** (ovvero **accuratezza**, **coerenza**, **completezza** e **attualità**), fra le 15 previste dallo Standard ISO/IEC 25012 come da indicazioni della Determinazione Commissariale n. 68/2013 di AgID.

Le suddette Linee Guida riportano un insieme esemplificativo di **misure** di queste caratteristiche di qualità, che si riportano qui a seguire.

Caratteristiche	Descrizione	Misure e funzioni di misura
Completezza	Il grado per cui il dato associato a un'entità presenta valori per tutti gli attributi attesi e relative istanze in un certo contesto.	Si individuano i seguenti livelli di completezza: 1. completezza di schema: percentuale di valori nulli per concetti e proprietà rispetto al numero totale di valori attesi; 2. completezza dei record: numero di dati elementari associati a un valore non nullo in un record, rispetto al numero di dati elementari del record per cui può essere misurata la completezza; 3. completezza di popolazione: percentuale di valori nulli rispetto a una popolazione di riferimento. Si noti che non sempre valori mancanti indicano incompletezza. Per esempio: si supponga di considerare dati relativi ai musei italiani e ai loro canali di contatto (telefono ed e-mail). Può capitare che i musei abbiano tutti un indirizzo e-mail ma non per tutti è presente un numero di telefono.
Accuratezza	Il grado in cui gli attributi rappresentano in maniera corretta il valore reale del dato in uno specifico contesto	Si individuano due tipi di accuratezza: 1. sintattica: ad esempio Merio invece che Mario 2. semantica: ad esempio nel caso in cui si utilizzi Marco Rossi intendendo invece un'altra persona per es., Mario Rossi Una misura dell'accuratezza è data dal rapporto tra gli attributi dei dati che hanno valori accurati sintatticamente/semanticamente sul numero di attributi

Copia

REGIONE LAZIO

Caratteristiche	Descrizione	Misure e funzioni di misura
		dei dati per i quali è richiesta accuratezza sintattica/semantica.
Coerenza	Il grado in cui gli attributi del dato non sono in contraddizione con altri dati in uno specifico contesto	Per poter valutare la coerenza una misura è quella che consente di identificare le violazioni di regole semantiche definite su alcuni elementi dei dati. Per esempio, se una persona è "patentata" non può essere possibile che la sua età sia "17 anni". Essa può essere calcolata come il rapporto tra il numero di attributi dei dati i cui valori sono semanticamente corretti nel dataset sul numero di attributi dei dati per i quali sono state definite delle regole semantiche. Altra misura consiste nel rapporto tra il numero di valori duplicati per ogni attributo della base dati e il numero totale degli elementi della base dati.
Tempestività	Il grado in cui gli attributi del dato sono al "giusto tempo" rispetto al contesto di riferimento	La metrica è basata sull'uso dei metadati che indicano quando il dato è stato aggiornato l'ultima volta. Sulla base di questi metadati, si distinguono poi: 1. dati con periodicità di aggiornamento nota: in questo caso è possibile calcolare la tempestività in maniera esatta identificando se la data di ultima modifica del dato rispetto al tempo di misurazione ricada nell'intervallo della frequenza di aggiornamento; 2. dati con periodicità di aggiornamento media: in questo caso è possibile calcolare la tempestività media con una percentuale di errore.

TABELLA 1 - CARATTERISTICHE DI QUALITÀ E RELATIVE MISURE

Copia

REGIONE LAZIO

8.3. Check-list Verifica limiti giuridici

Si riporta qui a seguire, la *check list* di verifica degli aspetti giuridici, presente nelle Linee Guida Open Data nazionali di AgID.

Aspetto	Domanda
Protezione dei dati personali	Nel processo di apertura di atti, documenti e/o dati è stata attentamente verificata l'assenza di informazioni riguardanti persone fisiche identificate o identificabili anche indirettamente, ai sensi dell'art. 4, n. 1) del GDPR[48], valutati anche gli ulteriori aspetti di cui ai punti riportati di seguito? Più nello specifico: I dati sono liberi da ogni informazione personale che possa identificare in modo diretto l'individuo (nome, cognome, indirizzo, codice fiscale, patente, telefono, e-mail, foto, descrizione fisica, ecc.)? I dati sono liberi da ogni informazione indiretta che possa identificare l'individuo (caratteristiche personali che possono identificare facilmente il soggetto)? I dati sono liberi da ogni informazione sensibile riconducibile all'individuo? I dati sono liberi da ogni informazione relativa al soggetto che, incrociata con dati comunemente reperibili nel web (ad es. Google maps, linked data, ecc.), possa facilmente identificare l'individuo? I dati sono liberi da ogni riferimento a profughi, protetti di giustizia, vittime di violenze o in ogni caso categorie protette? I servizi di ricerca sui dati sono tali da poter filtrare i dati in modo da ottenere un solo record geolocalizzato, che sia facilmente e direttamente riconducibile ad una persona fisica? Nel caso in cui è stata verificata la presenza di dati e informazioni personali, riconducibili in maniera diretta o indiretta a persone fisiche identificate o identificabili, si è provveduto a individuare idonee tecniche di anonimizzazione come

[48] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016

Copia

REGIONE LAZIO

Aspetto	Domanda
	indicato nel Parere 05/2014 del Gruppo di lavoro Articolo 29 sulle tecniche di anonimizzazione[49]? In un'eventuale decisione di rendere riutilizzabili dati personali oggetto di un obbligo di pubblicazione online previsto da un idoneo presupposto normativo ai sensi dell'art. 2-ter, commi 1 e 3, del D. Lgs. 196/2003 (ad es. ai sensi del D. Lgs. 33/2013) o di accogliere eventuali richieste di riutilizzo degli stessi da parte di terzi, è stato valutato se, per quali finalità ed entro quali limiti e condizioni siano eventualmente leciti e non incompatibili utilizzi ulteriori dei dati personali resi pubblici (con esclusione dei dati di cui agli artt. 9-10 del GDPR), sulla base del principio di limitazione della finalità? È stata effettuata la valutazione d'impatto in materia di protezione dei dati, ai sensi dell'art. 35 del GDPR, al fine di ridurre il rischio di perdere il controllo sulle medesime informazioni o di dover far fronte a richieste di risarcimento del danno da parte degli interessati? Nella già menzionata analisi di impatto sono state effettuate tutte le valutazioni (anche inerenti alla tipologia di licenza) si è provveduto a coinvolgere il Responsabile per la protezione dei dati personali, la cui nomina è obbligatoria in ogni PA e nei soggetti indicati dall'art. 37 del GDPR?
Proprietà intellettuale della sorgente	L'ente è proprietario dei dati, anche se non sono stati creati direttamente da suoi dipendenti? Sei sicuro di non usare dati per i quali vi è una licenza o un brevetto di terzi? Se i dati non sono del tuo ente, hai un accordo o una licenza che ti autorizzi a pubblicarli?
Licenza di rilascio **Limiti alla pubblicazione**	Se i dati non sono del tuo ente, hai un accordo o una licenza che ti autorizzi a pubblicarli? Hai verificato che i dati e i documenti non siano tra quelli esclusi dall'applicazione del decreto legislativo n. 36/2006?
Segretezza	Hai verificato se non vi siano motivi di ordine pubblico o di sicurezza nazionale, motivi legati al segreto d'ufficio o al segreto di stato o al segreto statistico che ti impediscono la pubblicazione dei dati?
Temporalizzazione	I dati sono soggetti per legge a restrizioni temporali di pubblicazione?

[49] Parere 06/2013 del Gruppo di lavoro Articolo 29 sul riutilizzo dei dati aperti e delle informazioni del settore pubblico

Copia

REGIONE LAZIO

Aspetto	Domanda
Trasparenza	I dati hanno dei divieti di legge o giurisprudenziali che impediscono la loro indicizzazione da parte di motori di ricerca? Fermi restando i limiti e le garanzie per il trattamento dei dati personali, i dati rientrano nella lista dell'allegato A del d.lgs. 33/2013 D-LGS-33-2013[50]?

Copia

[50] Decreto legislativo 14 marzo 2013, n. 33

REGIONE LAZIO

8.4. Raccomandazioni sulla qualità dei dati

A seguire si riportano i riferimenti ad alcune pubblicazioni di riferimento a tema.

Uno sono le **Linee Guida Open Data nazionali** e in particolare il capitolo "Formati aperti per dati e documenti". In questo, per ogni formato, è indicato un elenco di raccomandazioni.

Questo capitolo delle Linee Guida nazionali è ispirato al documento "***Data.europa.eu data quality guidelines***"[51] (sempre in PDF), pubblicato dal portale europeo sui dati aperti, a cui si invita a fare riferimento, perché ancora più ricco e dettagliato, con questi 4 capitoli principali:

- Raccomandazioni per fornire dati di alta qualità
- Raccomandazioni per la standardizzazione dei dati (con vocabolari controllati dell'UE) e l'arricchimento dei dati
- Raccomandazioni per la documentazione dei dati
- Raccomandazioni per migliorare il "livello di apertura"

Di questo documento EU, esiste anche una versione HTML, consultabile comodamente a questo indirizzo: https://dataeuropa.gitlab.io/data-provider-manual/data-quality/

[51] https://op.europa.eu/en/publication-detail/-/publication/b601d9cc-b3c0-11ec-9d96-01aa75ed71a1

Copia

REGIONE LAZIO

8.5. Glossario

Termine	Definizione
Dataset	Rappresenta - nello specifico standard DCAT (*Data Catalog Vocabulary*) - un insieme coerente di dati, pubblicato o curato come unità da un'organizzazione.
Leggibile meccanicamente (in inglese *machine readable*)	Dati in un formato processabile automaticamente dai *computer*, macchine e sistemi informatici, senza necessità di interventi manuali.
Opera	Per "opera", nel contesto delle licenze, si intende qualsiasi creazione dell'ingegno caratterizzata da un elemento di originalità e che riflette uno sforzo creativo o intellettuale
Primario (dataset)	Un "dataset primario" si riferisce a un insieme di dati raccolti direttamente alla fonte, senza essere stati modificati o elaborati da terze parti. Questi dati sono essenzialmente "grezzi" o nella loro forma originale
Ridistribuzione	La ridistribuzione si riferisce alla condivisione dei dati aperti con altre persone o organizzazioni.
Risorsa	La risorsa - nello specifico standard DCAT (*Data Catalog Vocabulary*) - è qualsiasi contenuto (file, servizi web, ecc.) incluso nel dataset. Nel contesto del profilo di metadatazione DCAT-AP, "distribuzione" è sinonimo di "risorsa".
Riutilizzo	Il riutilizzo si riferisce all'utilizzo dei dati pubblici per scopi diversi da quelli per cui sono stati originariamente creati o raccolti.
URI Dereferenziabile	È un URI che fornisce informazioni sufficienti per identificare in modo univoco una risorsa specifica o un oggetto all'interno di un sistema o di un'applicazione web e di accedere alla sua rappresentazione.

copia