

Piano nazionale di digitalizzazione del patrimonio culturale 2022-2023

**Linee guida per la redazione del piano di
gestione dei dati**

(DATA MANAGEMENT PLAN)



Versione in consultazione

Crediti

Il presente documento è stato prodotto dall'Istituto centrale per la digitalizzazione del patrimonio culturale – Digital Library, con il contributo di:

Luigi Cerullo, Tiziana Mancinelli, Daniele Metilli, Antonella Negri, Chiara Veninata.

Sommario

Introduzione.....	5
1. Cos'è il DMP e a che cosa serve.....	6
2. Il DMP nel Piano nazionale di digitalizzazione.....	8
3. Data governance interna.....	10
4. Attività preliminari.....	12
6. Scheda istituto.....	15
7. Scheda progetto.....	16
7.1. Informazioni amministrative.....	17
7.1.1. Identificazione del progetto.....	17
7.1.2. Ruoli e responsabilità ⓘ.....	17
7.2. Dati.....	19
7.2.1. Produzione e raccolta dei dati ⓘ.....	19
7.2.2. Tipologie e formati di dati ⓘ.....	19
7.2.3. Origine dei dati ⓘ.....	20
7.2.4. Tipologia di beni ⓘ.....	20
7.2.5. Migrazione di dati esistenti ⓘ.....	20
7.2.6. Quantità di dati prevista ⓘ.....	21
7.3. Metadati.....	22
7.3.1. Livello di metadattazione ⓘ.....	22
7.3.2. Metodologie per la metadattazione ⓘ.....	23
7.3.3. Standard di metadattazione ⓘ ©.....	23
7.4. Qualità dei dati e documentazione.....	24
7.4.1. Analisi della qualità dei dati ⓘ ⓘ.....	24
7.4.2. Bonifica dei dati ⓘ.....	25
7.4.3. Nomenclatura degli oggetti digitali ⓘ ©.....	25
7.5. Open data.....	27
7.5.1. Livello open data ⓘ.....	27
7.5.2. Identificatori ⓘ ⓘ.....	28

7.5.3. Standard per gli open data ⓘ ⓘ	28
7.5.4. Ontologie e vocabolari di riferimento ⓘ ⓘ	28
7.5.5. Licenze open data ⓘ ⓘ	29
7.5.6. Integrazione con altrimdataset ⓘ	29
7.6. Aspetti etici e legali	30
7.6.1. Dati personali e consenso informato ⓘ	30
7.6.2. Diritto d'autore ⓘ	30
7.6.3. Possibili fonti di <i>bias</i> ⓘ	31
7.7. Archiviazione, pubblicazione e preservazione dei dati	33
7.7.1. Archiviazione dei dati ⓘ	33
7.7.2. Pubblicazione dei dati ⓘ ⓘ	34
7.7.3. Preservazione dei dati ⓘ	36
7.7.4. Procedure di <i>backup</i> ⓘ ⓘ	36
7.7.5. Certificazione per la conservazione a lungo termine ⓘ	37
7.8. Sicurezza dei dati	38
7.8.1. Disposizioni per la sicurezza dei dati ⓘ	39
7.8.2. Standard per la sicurezza dei dati ⓘ	39
7.8.3. Rischi per la sicurezza dei dati ⓘ	40
7.8.4. Sicurezza dei dati sensibili ⓘ	40
8. Strumenti per la compilazione del DMP ⓘ	41
Normativa di riferimento	42
Allegato: FAQ per la pubblicazione dei dati aperti	44

Introduzione

Al pari di una qualsiasi attività che interessa il patrimonio culturale e i luoghi della cultura, anche i progetti di digitalizzazione, e più in generale la gestione di banche dati, necessitano di un'accurata programmazione che dia conto di tutte le fasi del processo generativo. Inoltre i dati, una volta prodotti, rappresentano un patrimonio che va correttamente gestito al fine della sua permanenza nel tempo.

Il piano di gestione dei dati (Data Management Plan, abbreviato DMP) è uno strumento che contribuisce a tale obiettivo, documentando le modalità di creazione, raccolta, gestione, conservazione e valorizzazione delle informazioni nell'ambito di un progetto di gestione dei dati o di digitalizzazione. Lo scopo della pianificazione della gestione dei dati e degli oggetti digitali è quello di garantire che essi siano tenuti al sicuro in tutte le fasi della filiera, dalla produzione all'archiviazione e/o pubblicazione, nonché di testimoniare e documentare tecnologie, metodologie e modalità impiegate nella definizione e nella realizzazione di un progetto digitale.

Questo approccio si estende alle diverse tipologie di dati prodotti dall'attività degli istituti culturali: dati provenienti da nuove o precedenti digitalizzazioni di beni culturali, prodotti della ricerca, documenti digitali di vario genere connessi a qualunque attività. Il presente documento, che correda il Piano nazionale di digitalizzazione del patrimonio culturale (PND), fornisce le linee guida per la redazione dei piani di gestione dei dati nello specifico ambito dei progetti di digitalizzazione e della gestione di banche dati sui beni culturali. È stato redatto avendo a riferimento le "Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico" elaborate da AgID - Agenzia per l'Italia Digitale¹ (di seguito anche "Linee guida AgID").

In questa prima versione, l'obiettivo del documento è introdurre la metodologia al fine di una sua concreta sperimentazione sul campo. In seguito, in relazione agli aggiornamenti del PND, saranno definiti gli strumenti operativi per rendere la compilazione dei DMP un metodo sistematico per presentare e accompagnare i progetti di digitalizzazione.

¹ Cfr. <https://www.dati.gov.it/linee-guida-valorizzazione-patrimonio-informativo-pubblico>.

1. Cos'è il DMP e a che cosa serve

I DMP sono strumenti che permettono di raccogliere informazioni su come i dati vengono acquisiti, descritti, elaborati, gestiti, definendone i piani per la loro condivisione e conservazione nel tempo, indicando eventuali restrizioni da applicare ai dati della cultura nelle varie fasi del progetto.

Il DMP delinea inoltre i principi e raccoglie indicazioni fondamentali su quali standard di metadati possono essere utilizzati, come i dati verranno archiviati e protetti dal rischio di perdite accidentali, come verranno condivisi e se esistono restrizioni sulla condivisione o sul riuso. Il DMP definisce inoltre ruoli e responsabilità durante l'intero ciclo di vita dei dati e, nei sistemi complessi, ha l'ulteriore finalità di ottimizzare la produzione dei dati e razionalizzarne la gestione, in ossequio al principio *once only* applicato alla creazione del dato².

Il piano di gestione dei dati consente agli istituti di pianificare, documentare fin dall'inizio e successivamente monitorare la gestione dei dati di un progetto di digitalizzazione, verificando poi passo dopo passo il rispetto di quanto è stato deciso.

Il DMP, dunque, aiuta a mettere a fuoco tutte le azioni che devono essere intraprese da chi avvia un progetto di digitalizzazione o di elaborazione di risorse digitali, affinché non vengano sottovalutati quegli aspetti che garantiscano la sostenibilità del processo nel tempo, quali ad esempio:

- il flusso di lavoro di tutto il ciclo di vita dei dati, dalla produzione alla pubblicazione;
- il rispetto degli obiettivi del finanziamento, il soddisfacimento delle *policy* di riuso e i vincoli di copyright;
- la preservazione dei dati a lungo termine, salvaguardando dal rischio di perdita il lavoro che spesso si articola in un intervallo temporale di più anni;
- l'utilizzo degli standard per la qualità e l'interoperabilità dei dati.

È un documento che garantisce chiarezza sia all'interno del progetto, sia all'esterno, ponendo le basi per la corretta fruizione e riuso dei dati prodotti. È prima di tutto un documento utile internamente, per la gestione del progetto stesso e dei relativi dati. Infatti, il DMP stabilisce chi ha la

²Il dato va prodotto una sola volta, deve essere conservato in un'unica posizione, ne va garantita la qualità e l'aggiornamento e tutti gli uffici del MiC lo utilizzano. Questo è particolarmente utile nella gestione delle cosiddette anagrafiche degli istituti culturali e dei beni culturali.

responsabilità di gestire i vari aspetti del progetto e quali sono i ruoli di ciascuno, in modo che tutti i partecipanti sappiano con certezza a chi fare riferimento per le varie attività che compongono un progetto di digitalizzazione o la conservazione di materiali digitali. Ciò è importante anche in prospettiva futura, nel caso fosse necessario recuperare queste informazioni nel tempo.

Il DMP consente inoltre di verificare e il rispetto dei requisiti giuridici e legali, considerare e prevenire eventuali rischi e criticità legate alla conservazione dei dati, e valutare le politiche di accesso e condivisione dei dati, tenendo traccia delle decisioni prese e delle soluzioni adottate. In questo modo, è possibile riconoscere e affrontare le problematiche tecniche e legali fin dall'inizio del progetto di ricerca (open source, database, licenze d'uso, tutela del diritto d'autore, ecc.).

È in ogni caso consigliabile aggiornare il DMP ogni volta che si ritiene necessario rivalutare le decisioni prese in precedenza. Nel corso del progetto possono infatti emergere nuovi requisiti che non erano stati previsti all'inizio. In questo senso, il DMP è un documento vivo e dinamico.

Concludendo, il DMP è uno strumento:

- *aperto*, perché accessibile e disponibile alla consultazione interna, ma potenzialmente anche esterna;
- *dinamico*, perché viene aggiornato periodicamente in base all'avanzamento del progetto;
- *condiviso*, perché è realizzato coinvolgendo tutti i diversi attori che prendono parte al processo di produzione e gestione dei dati.

2. Il DMP nel Piano nazionale di digitalizzazione

Il PND prevede che ciascun istituto, prima di intraprendere un progetto di digitalizzazione o la creazione di una banca dati, rediga il proprio Piano di gestione dei dati al fine di delineare come le risorse digitali saranno trattate in fase di acquisizione, processamento, conservazione e valorizzazione³.

Come previsto nel PND, il Ministero della Cultura (MiC) ha tre livelli di *governance* nella gestione dei dati:

- Direzioni generali: l'Istituto centrale per la digitalizzazione del patrimonio culturale (ICDP), in accordo con le Direzioni generali, definisce le *policy* per i dati del Ministero e si occupa della gestione dell'infrastruttura nazionale dei dati del patrimonio culturale (cfr. par. 2.1.a del PND), curando l'accesso e la conservazione a lungo termine degli oggetti digitali;
- Istituti centrali: raccolgono e gestiscono i dati a livello nazionale relativamente a ciascun dominio di competenza; ne curano la pubblicazione sui portali nazionali coerentemente con le proprie finalità istituzionali;
- Istituti periferici: producono, gestiscono e valorizzano i dati del patrimonio culturale per le proprie finalità istituzionali, utilizzando preferibilmente i software descrittivi e gestionali messi a disposizione gratuitamente dagli Istituti centrali e/o dalle Direzioni generali, ovvero - nel caso di utilizzo di sistemi locali - trattano i dati secondo gli standard nazionali e li rendono disponibili per le procedure di *data harvesting*⁴ svolte dall'ICDP per la condivisione con l'infrastruttura nazionale dei dati del patrimonio culturale;

Coerentemente con i tre livelli di *governance* descritti, e in relazione alle proprie caratteristiche e finalità, ciascun Istituto redige i DMP dei progetti di cui è responsabile:

³ Cfr. il paragrafo 2.2.a del Piano nazionale di digitalizzazione.

⁴ Il *data harvesting* è un processo in cui uno script viene utilizzato per estrarre automaticamente una grande quantità di dati di web site e/o di dataset per analizzarli per altri scopi.

- l'ICDP gestirà i dati del MiC a livello apicale per la realizzazione della Digital Library della cultura italiana e per l'alimentazione del repository di dati aperti dati.beniculturali.it;
- a livello centrale il DMP sarà compilato dalle Direzioni generali e dagli Istituti centrali in relazione alla gestione dei dati raccolti nei sistemi informativi di dominio;
- a livello periferico, gli istituti redigeranno il DMP per i progetti di digitalizzazione o di elaborazione dati di competenza.

Esistono diversi modelli a cui ispirarsi per predisporre un piano di gestione dei dati e nell'ambito di progetti di ricerca nazionali ed internazionali sono previsti *template* specifici o appositi *tool* da utilizzare prevalentemente online per la compilazione del piano. La struttura del DMP va sempre adattata al contesto specifico, ma le macro categorie delle informazioni richieste sono comunque comuni ai diversi modelli.

Il modello di Data Management Plan che viene proposto di seguito rappresenta una delle possibili modalità di formalizzazione delle informazioni necessarie allo scopo delineato; nei prossimi mesi l'ICDP avvierà una fase di sperimentazione alla fine della quale il modello potrà essere implementato in un *web tool* che verrà messo a disposizione degli Istituti.

3. Data governance interna

Al fine di una corretta redazione del DMP, è necessario che ogni Istituto della cultura definisca una chiara *data governance* interna, con l'individuazione di ruoli e responsabilità, che tenga conto dei processi consolidati e delle necessità dei diversi uffici. L'Agenzia per l'Italia Digitale (AgID) nelle già citate *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico*, delinea la struttura di un possibile gruppo di lavoro che opera in parte a livello centrale e in parte in ogni ufficio per gestire il processo di gestione dei dati. A seconda delle dimensioni degli uffici, alcune figure professionali possono coincidere o possono essere ulteriormente distinte, è quindi opportuno che ogni Istituto individui nel dettaglio i ruoli e le figure relativamente al suo specifico processo di gestione dei dati. Sulla base dell'organizzazione degli uffici del MiC, e contestualizzando la struttura della *governance* proposta dalle linee guida AgID, si elencano di seguito alcuni dei principali ruoli che andranno definiti prima dell'avvio delle attività di produzione/gestione di dati e risorse digitali:

- Responsabile del progetto: può avere un ruolo di coordinamento del progetto e/o di responsabilità scientifica.
- Responsabile della banca dati: è la figura che ha la responsabilità del procedimento amministrativo che popola la specifica fonte del dato, che ne cura la qualità e il relativo aggiornamento;
- Responsabile tecnico della banca dati: è una figura coordinata dal responsabile della banca dati con conoscenze informatiche e con un ruolo operativo sul sistema gestionale. Fornisce indicazioni circa il reperimento concreto dei dati dalla base dati, e cura il monitoraggio dei vari "connettori" che a partire dalla base dati espongono eventualmente il dato come dato aperto.
- Responsabile tematico della banca dati: si tratta di un esperto di dominio che conosce in modo approfondito l'Istituto e la storia dei dati su cui l'ufficio opera. Spesso propone nuovi dataset da esporre a partire dal sistema gestionale corrispondente e cura eventuali valutazioni di dominio o relative al significato dei dati. Ha anche la possibilità di compiere bonifiche e semplici adeguamenti sulla banca dati, su segnalazione o su valutazioni proprie. Riferisce invece al Responsabile della banca dati la necessità di eventuali variazioni strutturali al sistema gestionale che insiste sui dati.
- Responsabile Open Data (o Data Manager): conosce i dati dell'Istituto nel loro insieme, redige le procedure operative per lo scambio dati tra le diverse figure coinvolte e pianifica la strategia

di apertura e di diffusione dei dati, sia per quanto riguarda i dati di tipo più propriamente amministrativo, sia per i dati più tipicamente descrittivi sul patrimonio culturale. Si coordina con il Gruppo di lavoro Open Data, definito a livello degli istituti centrali del MiC.

- Curatore dei dati (o Digital Data Curator): è la figura che ha la responsabilità del processo creazione dei dati, dalla selezione dei più opportuni modelli di metadattazione, nonché delle procedure per la preservazione dei dati in vista dell'accesso e del riuso, anche sul lungo periodo.

Le figure sopra descritte dovranno poi coordinarsi con i seguenti gruppi di lavoro e uffici definiti ai livelli più alti della *governance*:

- Gruppo di lavoro Open Data: è definito a livello degli Istituti centrali; valuta le esigenze di pubblicazione dati e, nel rispetto della normativa, promuove l'uso e la diffusione degli Open Data. Si occupa di analisi della qualità dei dati, di definizione delle interfacce d'accesso ai dati, di promozione di applicazioni e servizi digitali sviluppate a partire dai dati pubblicati. Si occupa della formazione tecnica e concettuale all'interno dell'amministrazione sui temi legati al paradigma Open Data, anche sulla base delle linee guida pubblicate dall'Agenzia per l'Italia Digitale. Alcuni membri del *team* (e.g., esperti di tecnologie Web, esperti GIS, esperti di tecnologie e strumenti per i Linked Data) possono occuparsi della gestione del processo di apertura del dato dal punto di vista IT.
- Ufficio giuridico-amministrativo: fornisce consulenza su aspetti come la definizione delle licenze e delle note legali associate ai dati, sia su tutte quelle problematiche di tipo giuridico o amministrativo, comprese quelle di *privacy*, di finalità del dataset e di trattamento del dato personale ove presente. All'interno del MiC questo ruolo è svolto dalle Direzioni generali di settore e dall'ICDP.
- Ufficio statistica: l'ufficio statistica gioca un ruolo importante sia nel promuovere nuove tipologie di dataset da esporre, sia nel validare dal punto di vista metodologico e statistico i dati pubblicati e le loro visualizzazioni. All'interno del MiC questa funzione è svolta dall'Ufficio Statistica della Direzione Generale Bilancio.

Nel DMP possono poi essere individuati altri ruoli e figure specifiche che avranno la responsabilità di supervisionare la produzione e gestione dei dati del singolo progetto.

4. Attività preliminari

Prima della redazione del DMP, occorre che l'istituto abbia eseguito le attività preliminari previste dalle Linee guida AgID, e in particolare:

- Censimento dei dati. All'interno dei singoli uffici o dei vari settori dell'amministrazione è necessario condurre il censimento delle banche dati e dei progetti di digitalizzazione, anche per quelli in itinere o che riguardano il recupero di digitalizzazioni e dati pregressi. Dovranno essere raccolte informazioni in relazione alle caratteristiche dei dati archiviati o che verranno acquisiti, alla loro identificazione e descrizione. Ciascun soggetto preposto alla gestione di una particolare base di dati indica poi al responsabile Open Data, tra le altre cose, le caratteristiche descrittive del dato, i tracciati record, il tasso temporale di aggiornamento, e ogni altra informazione utile a far comprendere le caratteristiche peculiari dei dati. AgID raccomanda l'adozione di un approccio di tipo "*demand-driven*" per individuare i dati che tenga conto dell'impatto economico e sociale nonché del livello di interesse degli utilizzatori suddivisi opportunamente per categorie (ad esempio cittadini, imprese, altre pubbliche amministrazioni), dei loro requisiti e delle loro necessità.
- Analisi giuridica delle fonti. Alla fase di censimento fa seguito l'analisi giuridica delle fonti del dato, fondamentale per creare un servizio equilibrato nel rispetto della funzione pubblica e dei diritti dei singoli individui. Essa evidenzia limitazioni d'uso, determinazione dei diritti e dei termini di licenza. AgID fornisce una breve checklist⁵, utile per verificare se tutti gli aspetti giuridici siano stati valutati.
- Analisi delle politiche di accesso e licenza. 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. Esistono infatti casi in cui i dati possono essere diffusi solo in forma anonima, ossia a un livello di aggregazione tale da impedire di identificare le persone cui i dati si riferiscono, oppure per ragioni connesse alla sicurezza del bene a cui si riferiscono. A tal fine, è bene definire delle politiche di accesso ai dati in cui sia indicato un profilo di accesso specifico per ogni dato,

⁵ Cfr. voce specifica nelle FAQ allegate.

dettato dai diritti sull'informazione di base, dalle norme o dalle *policy* in atto.

- Sistema di conservazione e storicizzazione. I dataset rilasciati costituiscono non solo una risorsa per la collettività, ma un prezioso patrimonio anche per le pubbliche amministrazioni che possono in questo modo archiviare in modo alternativo i loro dati in modalità indipendente dagli applicativi software originali che li hanno prodotti. Per questo motivo è importante premunirsi di un sistema di archiviazione/conservazione che mantenga le diverse versioni dei dati nel lungo periodo. A tal fine si raccomanda di assicurare che le versioni stesse siano accessibili a un URL stabile, che sia anche documentato unitamente alla pubblicazione del dato.

Una volta eseguite queste attività, l'istituto disporrà delle informazioni necessarie per la corretta compilazione del DMP.

5. Struttura del Data Management Plan

Il modello di DMP che si propone si suddivide in una scheda istituto, che contiene le informazioni generali relative agli aspetti organizzativi dell'istituto e ai livelli di responsabilità, e in una o più schede progetto, che contengono le informazioni di dettaglio relative a ogni specifico progetto.

Data governance

La scheda progetto del DMP è compilata dall'Istituto che produce e gestisce i dati, con la collaborazione, ove opportuno, degli istituti centrali, delle direzioni generali e della *Digital Library*. Di seguito, è indicata con ⓘ la compilazione da parte dell'istituto, con © il contributo degli Istituti centrali o direzioni generali, e con @ il contributo della *Digital Library*.

6. Scheda istituto

Nella scheda istituto sono riportate le informazioni generali relative all'istituto e ai ruoli e alle responsabilità interne relative alla produzione e gestione dei dati previste dalle Linee guida AgID. Questa scheda contiene esclusivamente le informazioni amministrative relative all'istituto. Lo scopo principale della scheda è tenere traccia dei ruoli che sono stati assegnati a livello generale e non variano a seconda dello specifico progetto (cfr. cap. 3).

È possibile riutilizzare la medesima scheda istituto per più progetti. Per i dettagli sui vari ruoli, di norma, non è quindi necessario modificare la scheda all'avvio di ogni progetto, occorre però aggiornarla ogni qual volta vi siano modifiche nell'assegnazione dei ruoli.

7. Scheda progetto

Nella scheda progetto sono riportate tutte le informazioni di dettaglio relative al singolo progetto di creazione di contenuti digitali; ha la funzione di descrivere tutti gli aspetti del progetto, a partire dalle informazioni amministrative fino ad arrivare alle tipologie di dati e metadati, al rispetto dei requisiti di qualità, accessibilità, riuso, conservazione, sicurezza, e così via. Il modello di scheda progetto proposto dalla *Digital Library* è suddiviso nelle seguenti sezioni:

1. Sezione "Informazioni amministrative". Descrive il progetto, le persone e gli uffici che sono responsabili della gestione dei dati.
2. Sezione "Dati". Descrive le tipologie e i formati dei dati prodotti, raccolti e conservati nell'ambito del progetto.
3. Sezione "Metadati". Descrive gli standard e gli schemi di metadati a cui fa riferimento il progetto.
4. Sezione "Qualità dei dati e documentazione". Descrive le misure che vanno seguite al fine di assicurare la qualità dei dati e dei metadati a essi relativi.
5. Sezione "Open data e principi FAIR". Descrive le misure adottate per garantire l'accessibilità e il riuso dei dati.
6. Sezione "Aspetti etici e legali". Descrive gli aspetti etici derivanti dal progetto e in che modo vengono affrontati.
7. Sezione "Archiviazione e preservazione dei dati". Descrive le misure adottate per archiviare e preservare i dati.
8. Sezione "Sicurezza dei dati". Descrive le misure adottate per garantire la sicurezza dei dati e la protezione dei dati personali.

È cura di ogni istituto definire le modalità di gestione dei dati seguendo le linee guida fornite dal MiC, e assicurare la corretta applicazione del DMP.

Di seguito vengono fornite alcune indicazioni per ogni sezione della scheda progetto.

7.1. Informazioni amministrative

La prima sezione della scheda progetto contiene le informazioni generali e amministrative relative al progetto, che devono essere definite e rese accessibili per garantire la corretta gestione della banca dati. Definisce inoltre la *data governance* del progetto, attribuendo in modo chiaro le responsabilità sulla base delle Linee guida AgID⁶.

In alcuni casi, specie negli istituti più piccoli, è possibile che a una stessa persona vengano attribuite responsabilità multiple.

Di seguito sono elencate le principali informazioni amministrative che vanno descritte nel DMP, con l'indicazione della responsabilità nella compilazione.

7.1.1. Identificazione del progetto

Nome del progetto ⓘ: nome del progetto ed eventuale acronimo.

Codice progetto ⓘ: al progetto può essere assegnato un identificatore univoco, in modo da riconoscere con precisione l'identità del progetto ed evitare possibili ambiguità. Obiettivi del progetto ⓘ: descrizione sintetica degli obiettivi che il progetto intende raggiungere, evidenziando la coerenza con quanto previsto dal PND.

7.1.2. Ruoli e responsabilità ⓘ

Responsabilità del progetto ⓘ: indicazione dei nominativi delle persone che rivestono i vari ruoli definiti (responsabile del progetto, responsabile della banca dati, responsabile tecnico della banca dati, responsabile tematico della banca dati, responsabile open data, *digital data curator*).

Ulteriori ruoli di gestione dei dati ⓘ: indicazione dei nominativi degli ulteriori ruoli e responsabilità previsti nella gestione dei dati e metadati del progetto. In relazioni alle varie fasi in cui si articola il progetto, può inoltre essere utile descrivere le responsabilità dei vari ruoli in maniera sintetica tramite una matrice cosiddetta "RACI", come segue:

- Responsible (R): coloro che lavorano per eseguire un determinato compito. Esiste almeno un ruolo di responsabile.

⁶ Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico, "Aspetti organizzativi e qualità dei dati". Agenzia per l'Italia Digitale. 2017.
<https://docs.italia.it/italia/daf/lq-patrimonio-pubblico/it/stabile/aspettiorg.html>

- **Accountable (A):** il solo che può approvare il corretto completamento di un compito e che delega il lavoro ai responsabili. Può esistere un solo ruolo *accountable* per uno specifico compito.
- **Consulted (C):** coloro che possono essere consultati in quanto esperti di dominio e con i quali instaurare una comunicazione bidirezionale.
- **Informed (I):** coloro che devono essere tenuti aggiornati sui progressi del processo, spesso al termine dello stesso.

La Tabella 1 riporta un esempio di possibile matrice RACI. In generale, le responsabilità assegnate varieranno a seconda delle specificità del progetto.

	Responsabile progetto	Responsabile banca dati	Referente tecnico	Referente tematico	<i>Ulteriori ruoli</i>
Produzione/Raccolta	A/R	R	R	R	—
Trasformazione	A/R	R	R	C	—
Metadattazione	R	A/R	R	R	—
Open Data	R	A/R	R	R	—
Pubblicazione	A/R	C	I	I	—

Tab. 1 – Esempio di matrice dei ruoli e responsabilità (RACI)

7.2. Dati

Lo scopo di questa sezione è documentare tutti gli aspetti relativi ai dati, consentendo di tenerne traccia nel corso del progetto. Vanno quindi descritte sinteticamente le tipologie di dati raccolte nel progetto o i documenti digitali conservati dall'Istituto, i formati utilizzati, l'origine dei dati e l'eventuale migrazione di dati esistenti. Contiene inoltre informazioni sintetiche sulla tipologia di documenti digitali o di beni oggetto di digitalizzazione, e sulla quantità di dati che si prevede di raccogliere nel corso del progetto.

Tutte queste informazioni sono utili per documentare i processi che generano e trasformano i dati del progetto. Di seguito sono descritti più nel dettaglio i principali aspetti relativi ai dati che è opportuno descrivere nel DMP. Di fianco al nome di ogni sezione è indicata la responsabilità nella compilazione.

7.2.1. Produzione e raccolta dei dati ⓘ

In questa sezione è opportuno descrivere le modalità di produzione e raccolta dei dati in relazione agli obiettivi del progetto. In particolare, occorre descrivere dettagliatamente in che modo i dati vengono raccolti/prodotti all'origine, chi è responsabile della raccolta o della produzione, se la raccolta/produzione avviene in maniera automatica o manuale.

In caso di produzione automatica dei dati, sarebbe rilevante esplicitare quali strumenti software vengono utilizzati per questa attività e se e in che modo i dati vengono trasformati dopo la raccolta. Nel caso vengano trasformati, quali strumenti software vengono adottati. Per ulteriori informazioni, fare riferimento al capitolo I, sezione D delle *Linee guida per la digitalizzazione del patrimonio culturale* e alle *Linee guida per l'acquisizione, la circolazione e il riuso delle riproduzioni dei beni culturali in ambiente digitale*.

7.2.2. Tipologie e formati di dati ⓘ

È opportuno che il DMP descriva tutte le tipologie di dati prodotti/raccolti nel progetto, e per ogni tipologia il formato in cui sono raccolti i dati, facendo riferimento al capitolo II delle *Linee guida per la digitalizzazione del patrimonio culturale* e alle linee guida AgID sui formati aperti per i dati e i

documenti⁷. L'uso dei formati consigliati dalle linee guida garantisce la migliore esecuzione del processo di digitalizzazione e della valorizzazione del bene. Inoltre, buona pratica per la conservazione a lungo termine di progetti di digitalizzazione, è la generazione di file master (*lossless* e non compresso) e uno o più file derivati.

7.2.3. Origine dei dati ⓘ

Un aspetto che va sicuramente monitorato e tracciato è l'origine dei dati che ci si trova a trattare. Affinché possano essere documentate la conoscenza delle origini e gli eventuali spostamenti dei dati, è importante indicare chi ha originariamente prodotto i dati e in che modo, quando e dove sono stati prodotti i dati, e se si tratta di dati nuovi o già esistenti. Inoltre, se si tratta di dati esistenti, è opportuno indicarne la provenienza e descriverne le modalità di migrazione (sezione 7.2.5).

7.2.4. Tipologia di beni ⓘ

Per ciò che riguarda un progetto di digitalizzazione, la precisazione della tipologia dei beni a cui si riferiscono gli oggetti digitali può essere di supporto alla documentazione e a una successiva verifica. In particolare, è utile descrivere nel dettaglio tutte le tipologie di beni a cui si riferiscono i dati e per ogni tipologia indicare la quantità di beni oggetto di digitalizzazione, l'intervallo temporale e il luogo geografico di origine dei beni. Queste informazioni potranno essere utili anche a fini statistici e di ricerca.

7.2.5. Migrazione di dati esistenti ⓘ

Nel caso di migrazione o conversione di dati esistenti, è opportuno indicare nel DMP le tipologie di migrazione e conversione e i metodi utilizzati, verificando se ci siano incongruenze tra il vecchio e il nuovo formato, e controllando che non siano più utilizzati formati obsoleti. A corredo di queste informazioni va inoltre affiancata una descrizione dei vari passaggi della migrazione e della conversione, con particolare riferimento alla pulizia dei dati, alla preparazione e rimozione di informazioni duplicate per garantire l'integrità dei dati, alla qualità e alla sicurezza dei dati. In questo modo, sarà possibile tenere traccia delle procedure seguite ed eventualmente ripeterle in caso di necessità. Inoltre, nel DMP è opportuno indicare quali strumenti

⁷ <https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/arch.html#formati-aperti-per-i-dati-e-documenti>

software vengono utilizzati per compiere queste operazioni e descrivere l'archivio dei file sorgenti.

7.2.6. Quantità di dati prevista ⓘ

È importante stimare fin dall'inizio del progetto la quantità di dati che si prevede di raccogliere, o almeno il loro ordine di grandezza. In base alla quantità di dati indicata, potranno essere previsti diversi profili di configurazione, con riferimento anche al conferimento all'infrastruttura software per i beni culturali.

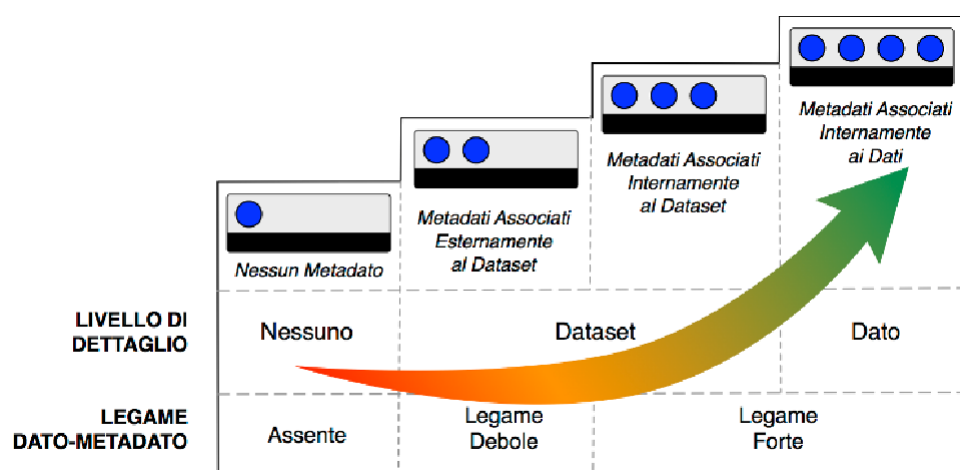
7.3. Metadati

I metadati sono informazioni strutturate che descrivono set di dati attraverso vocabolari controllati condivisi da una comunità per domini di conoscenza specifici; supportano il modello, la comprensione e la gestione di altri dati e informazioni. La descrizione dei dati e la documentazione dei tipi di metadati utilizzati (amministrativi, semantici, descrittivi) deve essere accompagnata dalla documentazione e da riferimenti precisi agli standard e alle metodologie applicate per i vocabolari di dominio a sostegno dell'interoperabilità e delle normative di riutilizzo.

In questo modo, il DMP documenterà tutte le decisioni relative ai metadati del progetto, consentendo di rendere i dati riutilizzabili in futuro, sia da chi li ha prodotti sia da chi li riusa. Di seguito sono segnalati gli aspetti relativi ai metadati che vanno descritti nel DMP. Di fianco al nome di ogni campo è indicata la responsabilità nella compilazione.

7.3.1. Livello di metadattazione

Le linee guida AgID prevedono quattro possibili livelli di metadattazione⁸, come descritto nella Fig. 1. Si consiglia di adottare il livello più alto che è tecnicamente possibile implementare. Il primo livello non include metadati, ed è quindi fortemente sconsigliato. Il secondo e il terzo livello hanno esclusivamente metadati associati al dataset, mentre il quarto livello prevede metadati associati internamente al dato stesso.



⁸ <https://docs.italia.it/italia/daf/ig-patrimonio-pubblico/it/stabile/modellometadati.html#livelli-del-modello-per-i-metadati>

Fig. 1 – Livelli di metadattazione (fonte: <https://docs.italia.it/italia/daf/lq-patrimonio-pubblico/it/stabile/images/Metadati.png>)

7.3.2. Metodologie per la metadattazione ⓘ

Nel DMP sono descritte le metodologie di metadattazione, specificando in che modo vengono prodotti i metadati e chi ne è responsabile. Per ulteriori informazioni si rimanda a quanto contenuto nelle *Linee guida per la digitalizzazione del patrimonio culturale* e ai regolamenti o prassi in vigore nella comunità scientifica di riferimento.

7.3.3. Standard di metadattazione ⓘ ©

Il risultato delle precedenti tappe del modello operativo si traduce nella produzione di metadati che certificano le caratteristiche del dato. Come detto precedentemente, la metadattazione, sia dei singoli dati che degli interi dataset, è cruciale: una delle problematiche più diffuse relative ai dati del MiC è la molteplicità di copie disponibili di una stessa informazione, senza che sussista la necessaria certezza sulle caratteristiche e sulla validazione di ciascun rilascio.

Nel DMP è opportuno descrivere tutti gli standard di riferimento che sono stati adottati per la metadattazione, facendo riferimento al capitolo IV delle *Linee guida per la digitalizzazione del patrimonio culturale*. In particolare, per i metadati è opportuno seguire lo standard METS (*Metadata Encoding and Transmission Standard*), che codifica i metadati descrittivi, amministrativi e strutturali riguardanti le risorse digitali.

7.4. Qualità dei dati e documentazione

Questa sezione descrive gli approcci per assicurare la qualità dei dati (convenzioni di denominazione dei file, revisione, vocabolari controllati, regole di validazione/verifica dei dati). Il controllo qualità sarà previsto dopo l'inserimento dei dati per convalidare e revisionare l'inserimento dei dati e dei loro metadati.

Di seguito sono indicati gli aspetti relativi alla qualità dei dati che è opportuno descrivere nel DMP, e per ognuno è indicata la responsabilità nella compilazione.

7.4.1. Analisi della qualità dei dati

Secondo la norma ISO 9000:2015, la qualità è la totalità degli elementi e delle caratteristiche di un prodotto o servizio che concorrono alla capacità dello stesso di soddisfare esigenze espresse o implicite. Le misure per quantificare la qualità dei dati sono individuate dallo standard ISO/IEC 25012:2008, divenuto norma italiana UNI ISO/IEC 25012:2014, che definisce un insieme di 15 caratteristiche specifiche per la caratterizzazione della qualità dei dati: accuratezza, attualità, coerenza, completezza, credibilità, accessibilità, comprensibilità, conformità, efficienza, precisione, riservatezza, tracciabilità, disponibilità, portabilità e ripristinabilità. L'ISO/IEC 25024 estende l'ISO/IEC 25012 definendo 63 misure di qualità applicabili alle caratteristiche di qualità dei dati.

Le caratteristiche più rilevanti ai fini della redazione del DMP sono quattro: accuratezza, coerenza, completezza e attualità. La Determinazione Commissariale dell'Agenzia per l'Italia Digitale n. 68/2013 riporta una tabella con un insieme esemplificativo di misure, con riferimento allo standard ISO, a supporto delle attività di valutazione della qualità dei dati delle amministrazioni⁹. Nel DMP è opportuno descrivere le modalità di analisi della qualità dei dati adottate nel progetto, e in particolare, quale standard è stato adottato per valutare la qualità dei dati; quali misure di qualità dei dati sono in uso e per ogni misura, in che modo viene eseguita la misurazione. In caso di tipologie di dati multiple, queste informazioni devono essere riportate per ogni specifica tipologia.

⁹ <https://docs.italia.it/italia/daf/Ig-patrimonio-pubblico/it/stabile/aspettiorg.html#qualita-dei-dati>

7.4.2. Bonifica dei dati ⓘ

Generalmente l'analisi della qualità del dato può richiedere una fase di bonifica: i dati all'interno dei sistemi informativi o degli archivi di un'amministrazione possono talvolta essere "sporchi" e non rispondenti ai requisiti di qualità previsti dagli standard adottati nel progetto. La bonifica può essere basata sui dati oppure basata sui processi. Nel primo caso, il dataset viene corretto tramite un confronto con il mondo reale (anche con attività economicamente onerose come contattare direttamente i soggetti preposti alla gestione della base dati) e/o tramite un confronto incrociato (*matching*) con altri dataset.

La bonifica basata sui processi invece analizza le cause che hanno portato alla scarsa qualità del dato e rivede i processi di produzione del dato per garantirne la qualità nel tempo. Per esempio, se si riscontra che la scarsa accuratezza di una base di dati deriva da un processo di inserimento dati manuale, si può intervenire sulla fase di acquisizione introducendo metodi per minimizzare la possibilità di errori. La bonifica basata sui processi ha dunque il consistente vantaggio di essere una strategia risolutiva. Nel DMP è opportuno descrivere le attività di bonifica dei dati previste nel progetto, e in particolare:

- quali sono le attività oggetto di bonifica (es. inserimento dati)
- se la bonifica è basata sui dati oppure sui processi
- in caso di bonifica basata sui dati, quali sono le modalità di bonifica e quali sono i dataset di riferimento per la bonifica
- in caso di bonifica basata sui processi, quali interventi sono previsti per minimizzare gli errori
- quali software sono utilizzati per la bonifica

In caso di tipologie di dati multiple, queste informazioni devono essere riportate per ogni specifica tipologia.

7.4.3. Nomenclatura degli oggetti digitali ⓘ ©

Le convenzioni per nominare i file sono importanti per descrivere ciò che contengono e come si relazionano ad altri file. Lo sviluppo di una struttura è fatto attraverso l'identificazione degli elementi chiave del progetto, le differenze importanti e i punti in comune tra i file che l'istituto produce o gestisce. Questi elementi potrebbero includere aspetti come la data di creazione, il nome dell'autore, il nome del progetto, il nome di una sezione o una sottosezione del progetto, la versione del file, ecc. Un vantaggio nell'usare nomi di file unici e standardizzati è la capacità di seguire i nomi dei percorsi e collegarsi ad altri sistemi che richiedono nomi di file unici. Le

indicazioni per la nomenclatura degli oggetti digitali sono riportate nel Capitolo VI delle *Linee guida per la digitalizzazione del patrimonio culturale*. Nel DMP è opportuno descrivere la metodologia utilizzata per assegnare i nomi dei file, spiegando in che modo è implementata e quali software sono utilizzati per assegnare i nomi dei file. Inoltre, è importante indicare la modalità di costruzione del codice oggetto (che costituisce la parte variabile del nome) e la provenienza degli identificativi presenti nel codice oggetto (es. SBN). In caso di metodologie di nomenclatura multiple, sarà opportuno riportare queste informazioni per ogni specifica metodologia adottata.

7.5. Open data

La pubblicazione dei dati in formato aperto è una delle fasi del ciclo di vita dell'informazione pubblica. Nella redazione del DMP, occorre seguire le strategie per la produzione, pubblicazione e conservazione di dati di qualità definite dal Ministero e gli obiettivi previsti per il prossimo triennio.

La strategia del Ministero espressa nel PND, in coerenza con la normativa nazionale e con gli orientamenti generali della pubblica amministrazione, prevede che i dati raccolti siano aperti *by default*; qualora un *dataset* non possa essere pubblicato in formato aperto, è necessario motivare le ragioni ostative. In questa sezione vengono descritti gli standard adottati per gli open data, la licenza selezionata per la pubblicazione dei dati e il rispetto dei principi FAIR (*Findable, Accessible, Interoperable, Reusable*)¹⁰ che garantiscono la reperibilità, accessibilità, interoperabilità e riutilizzabilità dei dati. I *dataset* del Ministero saranno metadati e descritti secondo le indicazioni e i profili applicativi forniti da AgID e resi disponibili in un registro dei metadati appositamente costruito.

Il MiC prevede la pubblicazione dei dati aperti nel portale <https://dati.beniculturali.it>, che costituisce il repository dei dati aperti del Ministero. I dati pubblicati nel portale saranno automaticamente conferiti al portale nazionale dei dati aperti (dati.gov.it) e da qui all'equivalente portale europeo.

Di seguito sono indicati gli aspetti relativi agli open data che è opportuno descrivere nel DMP, e quali sono i responsabili della compilazione. Per una specifica di tutti i termini utilizzati si rimanda alle *FAQ per la pubblicazione di dati aperti* in calce al presente documento.

7.5.1. Livello open data ⓘ

Le Linee guida AgID, che riprendono la definizione di Open Data del W3C¹¹, prevedono cinque possibili livelli di open data, come mostrato nella voce corrispondente delle FAQ.

Il MiC pubblicherà i dati in formato aperto almeno di livello 3* mirando, nel tempo, a pubblicare solo dati di livello a 4* e 5* (*Linked Open Data*). Si

¹⁰Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

¹¹ <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html#x5-star-linked-open-data>

consiglia in ogni caso di adottare il livello più alto che è tecnicamente possibile implementare.

Nel DMP deve essere indicato se i dati saranno pubblicati in formato aperto e a che livello.

7.5.2. Identificatori

In caso di *Open Data* di livello 4* o 5*, è opportuno indicare nel DMP le metodologie adottate per la costruzione di URI (*Uniform Resource Identifier*) o IRI (*Internationalized Resource Identifier*) univoci e persistenti, facendo riferimento alle linee guida del W3C¹².

Nel DMP è importante descrivere in che modo vengono assegnati questi identificatori, come vengono costruiti e qual è il nome di dominio di riferimento.

Nel caso che il sistema locale adotti un modello “integrato” con l’infrastruttura dati nazionale (cfr. par. 7.7.2 del presente documento e par. 2.1.a del PND) gli URI verranno definiti nell’infrastruttura stessa; nel caso dell’adozione di un modello “federato” gli URI dovranno essere indicati dal singolo istituto.

7.5.3. Standard per gli open data

Un altro aspetto importante che va descritto nel DMP riguarda gli standard adottati per la produzione degli open data; va quindi registrato nel DMP quale versione è stata adottata di quale standard e per quale tipologia di dati viene utilizzato.

7.5.4. Ontologie e vocabolari di riferimento

Allo stesso modo, è opportuno documentare nel DMP le ontologie e vocabolari di riferimento adottati nel progetto, descrivendo le finalità per cui verranno utilizzati. Si raccomanda di modellare i dati sulla base dei vocabolari e delle ontologie di OntoPiA (cfr. voce specifica nelle FAQ allegate) in larga parte allineati (collegati) a standard aperti del Web e disponibili in formati aperti standard sulla piattaforma <https://github.com/italia/>. Gli uffici del MiC, per il tramite degli Istituti centrali, sono incoraggiati ad avviare un processo di standardizzazione sia per la rappresentazione di dati ricorrenti,

¹² <https://www.w3.org/TR/Id-bp/>

indipendenti dallo specifico dominio applicativo, come per esempio i dati sulle persone, sulle organizzazioni pubbliche e private, sui luoghi e gli indirizzi usando le ontologie di OntoPiA. In particolare, nel DMP andranno elencati le specifiche ontologie e vocabolari adottati, quale versione di ciascuno è stata utilizzata e per quale finalità. Nel caso si scegliesse di adottare ontologie o vocabolari non standard in alternativa a standard di dominio, è utile descrivere il perché di questa scelta. Inoltre, se nel corso del progetto vengono prodotti delle ontologie o dei vocabolari, è opportuno descriverli nel DMP, in modo da facilitarne il più possibile il riuso.

7.5.5. Licenze open data ⓘ ⓘ

La scelta di una licenza è indispensabile per determinare come poter riutilizzare il dataset. È quindi opportuno indicare nel DMP il nome e la versione delle licenze aperte adottate nel progetto, specificando a quali dati si applicano e fornendo anche un riferimento al testo completo della licenza (cfr. voce specifica nelle FAQ allegate).

7.5.6. Integrazione con altri dataset ⓘ

Uno dei principali vantaggi dell'adozione degli *open data* e dell'uso degli standard del Semantic Web è la facilità di integrazione o *linking* del dataset che viene prodotto con altri dataset già esistenti che seguono i medesimi standard. Ciò costituisce un valore aggiunto per il progetto stesso, che potrà riutilizzare e integrare i dati prodotti in precedenza, e al tempo stesso facilita l'interoperabilità e il futuro riuso dei nuovi dati che vengono prodotti. Esistono oggi molteplici basi di conoscenza che sono facilmente integrabili con un nuovo dataset. Ciò può essere fatto riutilizzando direttamente gli URI esterni, oppure importando i dati (nel rispetto delle rispettive licenze), o ancora eseguendo un *linking* tramite il linguaggio OWL (proprietà *owl:sameAs*). Infine, può essere anche prevista la ripubblicazione diretta dei dati prodotti nel progetto in una base di conoscenza esistente, andando ad arricchirla. Sarebbe opportuno ai fini anche dell'interoperabilità se è prevista l'integrazione con dataset esistenti, e per ognuno di questi dataset, qual è la tipologia di integrazione (es. riuso, *linking*, pubblicazione nella base dati esistente). In caso di *linking* o riuso diretto di URI, è utile indicare quali sono le modalità con cui avviene il *linking* o riuso, e se si tratta di un riuso diretto, quali sono le modalità di importazione dei dati e in che modo è stata valutata la compatibilità delle licenze. Infine, in caso di pubblicazione dei dati prodotti nel progetto in una base di conoscenza esistente, è opportuno descrivere in che modo è prevista la pubblicazione e in che modo è stata valutata la compatibilità delle licenze.

7.6. Aspetti etici e legali

Nella gestione di un progetto di digitalizzazione, vanno necessariamente affrontate alcune questioni etiche e di *policy* dei dati, che in taluni casi possono avere anche risvolti legali. Questa sezione descrive le modalità con cui vengono questi affrontati nella gestione dei dati.

7.6.1. Dati personali e consenso informato ⓘ

Il diritto alla protezione dei dati è sancito dalla Carta dei diritti fondamentali dell'UE e dal Trattato sul funzionamento dell'Unione europea, che rendono effettivo il diritto alla *privacy* delle persone fornendo loro il controllo sul modo in cui le informazioni che li riguardano vengono raccolte e utilizzate.

Le categorie di dati personali che meritano speciale protezione sotto il profilo della riservatezza sono definite dal Regolamento UE 2016/679 relativo alla protezione dei dati personali (GDPR). In particolare, i dataset pubblicati non devono contenere "dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona" (art. 9 del GDPR) e i "dati personali relativi alle condanne penali e ai reati o a connesse misure di sicurezza" (art. 10 del GDPR).

In questo paragrafo occorre dunque indicare:

- Se nel progetto verranno trattati dati personali, e se sì quali
- Chi è responsabile del trattamento dei dati personali
- Quali soggetti sono proprietari dei dati personali
- Se sono previste modalità di raccolta consenso informato, e se sì quali sono.

7.6.2. Diritto d'autore ⓘ

Il diritto d'autore in Italia è protetto dalla legge 22 aprile 1941 n. 633 (*Protezione del diritto d'autore e di altri diritti connessi al suo esercizio*), che descrive i diritti riconosciuti agli autori di un'opera, e quali sono le modalità di fruizione di tali diritti. La legge distingue il diritto morale, che è inalienabile e indipendente dall'utilizzazione economica dell'opera, dai diritti patrimoniali (pubblicazione, riproduzione, trascrizione, esecuzione e rappresentazione,

comunicazione, distribuzione, modificazione, traduzione, noleggio e prestito), che sono invece rinunciabili e soggetti a un limite temporale. In Italia, la durata prevista per i diritti patrimoniali è nella maggior parte dei casi di 70 anni.

I diritti relativi alle banche dati, e i doveri a cui sono soggetti gli utenti delle stesse, sono normati dagli articoli 102-bis e 102-ter della legge sul diritto d'autore, che prevedono una serie di vincoli e un diritto esclusivo sulla banca dati da parte del costituente, che ha una durata di 15 anni.

In un progetto di digitalizzazione, è fondamentale tenere in considerazione tutti gli aspetti del diritto d'autore (o *copyright*) e assicurarsi di rispettare la legge, in particolare nel momento in cui viene adottata una specifica licenza, che deve pertanto essere compatibile con i diritti che l'istituto detiene.

In questo campo vanno descritti gli aspetti relativi al diritto d'autore che vanno affrontati nel progetto, e in particolare:

- Indicare se i beni oggetto di digitalizzazione sono attualmente protetti dal diritto d'autore o da altri diritti di natura patrimoniale
- Se sì, indicare le tipologie di beni e la durata dei relativi diritti
- Se sì, indicare come si prevede di garantire il rispetto di tali diritti
- Indicare se i dati prodotti o raccolti nel progetto sono protetti dal diritto d'autore o da altri diritti di natura patrimoniale
- Se sì, indicare come si prevede di garantire il rispetto di tali diritti
- Indicare se gli eventuali dataset riutilizzati nel progetto (di cui al paragrafo 7.5.6) sono protetti dal diritto d'autore o da altri diritti di natura patrimoniale
- Se sì, indicare come si prevede di garantire il rispetto di tali diritti.

7.6.3. Possibili fonti di *bias* ⓘ

Nella produzione, modellazione ed elaborazione dei dati è possibile che vengano introdotti *bias*, ovvero distorsioni che rendono i dati non rappresentativi della realtà, in particolare per quanto riguarda la rappresentazione delle minoranze e delle comunità marginalizzate. È opportuno che nella gestione di un progetto siano individuate tutte le possibili fonti di *bias*, e che si prevedano meccanismi per correggerli (oppure, ove non fosse possibile, contestualizzarli).

La pubblicazione di dataset contenenti *bias* corre infatti il rischio di perpetuare le disuguaglianze sociali, ad esempio quelle di genere¹³, etnia, lingua, religione, orientamento sessuale. Nella rappresentazione di dati

¹³ D'Ignazio, C. and Klein, L. F. (2020). Data Feminism. Ideas Series. MIT Press.

storici, è inoltre particolarmente frequente che esistano *bias* "per omissione", ovvero che nel dataset siano state privilegiate talune categorie sociali rispetto ad altre al punto di rendere queste ultime invisibili¹⁴.

Qualora i dati fossero stati elaborati tramite metodi di intelligenza artificiale (es. apprendimento automatico di conoscenza da testi o immagini), occorre considerare i limiti degli attuali sistemi di *machine learning* e valutare attentamente i possibili *bias* che ne derivano¹⁵.

In questo paragrafo vanno descritti:

- I possibili *bias* derivanti dalla produzione e raccolta dei dati
- I possibili *bias* derivanti dal riuso di dati esistenti
- I possibili *bias* derivanti dai modelli utilizzati per rappresentare i dati
- I possibili *bias* derivanti dall'uso di sistemi di intelligenza artificiale
- Eventuali ulteriori tipologie di *bias* individuabili nel progetto
- Per ogni tipologia di cui sopra, le modalità con cui si prevede di farvi fronte

¹⁴ Ortolja-Baird, A. & Nyhan, J. (2021). "Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive." *Digital Scholarship in the Humanities*.

¹⁵ Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021.

7.7. Archiviazione, pubblicazione e preservazione dei dati

Questa sezione descrive le procedure per l'archiviazione, la pubblicazione e la preservazione a lungo termine dei dati del progetto.

7.7.1. Archiviazione dei dati ⓘ

Nel DMP sono descritte le modalità con cui sono archiviati i dati del progetto. È sempre consigliabile archiviare i dati in repository istituzionali. Il MiC dispone di un Centro Elaborazione Dati (CED) che si occupa dei servizi di *housing* e *hosting* della maggior parte dei sistemi informativi, dei siti e dei portali del MiC. In coerenza con la Strategia per la crescita digitale del Paese e il Piano Triennale per l'informatica nella PA¹⁶, AGID ha delineato una strategia *Cloud*¹⁷ che prevede un percorso di qualificazione per i soggetti pubblici e privati che intendono fornire servizi *cloud* alla Pubblica amministrazione, affinché queste ultime possano adottare servizi e infrastrutture di cloud computing omogenei, che rispettino elevati standard di sicurezza, efficienza e affidabilità. Inoltre, più recentemente, il Dipartimento per la trasformazione digitale con l'Agenzia per la cybersicurezza nazionale ha elaborato la Strategia Cloud Italia¹⁸ contenente gli indirizzi strategici per il percorso di migrazione verso il cloud di dati e servizi digitali della Pubblica Amministrazione.

Coerentemente con questi orientamenti, per garantire l'archiviazione e la preservazione nel tempo dei dati del patrimonio culturale, il Ministero nell'ambito dell'investimento M1C3 1.1 "*Digital Strategy and Platforms for Cultural Heritage*" del Piano Nazionale di Ripresa e Resilienza (PNRR) ha previsto di sviluppare un'infrastruttura (hardware e software) per la gestione delle risorse digitali e per l'orchestrazione di servizi e procedure, finalizzata alla raccolta, conservazione, organizzazione e distribuzione delle risorse digitali della cultura, in linea con le strategie europee e con le indicazioni del piano triennale dell'informatica nella PA. L'infrastruttura software (cfr. par. 2.1.a del PND) costituirà il nucleo centrale di una rete di sistemi, anche esterni, la cui interconnessione aggiunge valore ai diversi sistemi singolarmente considerati. Allo stesso modo l'infrastruttura consentirà l'interconnessione trasversale dei contenuti delle diverse collezioni digitali,

¹⁶ <https://docs.italia.it/italia/piano-triennale-ict/pianotriennale-ict-doc/it/2020-2022/index.html>

¹⁷ <https://www.agid.gov.it/it/infrastrutture/cloud-pa>

¹⁸ <https://innovazione.gov.it/dipartimento/focus/strategia-cloud-italia/>

superando la logica “a silos” o “ad aggregazione di silos”, che ha caratterizzato ad oggi la digitalizzazione delle collezioni digitali e la loro interconnessione. A questo scopo verranno integrate nell’infrastruttura le logiche più evolute del Web Semantico e del Web 2.0, ma anche le più promettenti tecnologie di intelligenza artificiale applicate alla multimedialità.

Si tratterà di un’infrastruttura aperta, congruente con le Linee guida per la digitalizzazione, adottando quanto verrà man mano stabilito nel PND. L’infrastruttura metterà a disposizione diversi strumenti di alimentazione del repository: le collezioni digitali che saranno contenute nel repository saranno esposte nella Digital Library della cultura, che ne costituirà il principale portale di accesso, ma anche di altri sistemi di accesso che intendano utilizzarne i servizi; si tratterà quindi di un repository aperto anche nella distribuzione delle collezioni digitali conservate, fatti salvi i vincoli sulla proprietà intellettuale.

È opportuno pertanto che gli Istituti, nella gestione dell’archiviazione dei dati del progetto tengano conto di quanto sopra descritto e si allineino alle indicazioni che verranno fornite dall’ICDP e dallo stesso MiC.

In relazione alle specifiche informazioni sull’archiviazione dei dati del progetto, è opportuno comunque definire:

- la localizzazione (fisica o virtuale) dei dataset
- le modalità di conferimento dei dataset
- le modalità di conservazione dei dataset
- il responsabile della gestione dei dataset

7.7.2 Pubblicazione dei dati

Se l’istituto ha - o prevede di realizzare - un’interfaccia utente di fruizione delle informazioni (portale o sito web), in questa sezione è importante indicare dove e come verranno pubblicati i dati e descrivere in che modo è garantito il rispetto delle raccomandazioni AgID per la pubblicazione dei dati¹⁹; nella pubblicazione occorrerà in particolare prestare attenzione ai seguenti aspetti:

- assegnare ai dataset nomi auto-esplicativi e fornire descrizioni testuali degli stessi;
- mettere in evidenza la licenza in uso;
- fornire strumenti di ricerca e navigazione dei dati;
- fornire statistiche di uso, accesso e produzione, notifiche di aggiornamento dei dataset e del sito web;

¹⁹ <https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/publdatigov.html>

- fornire strumenti per semplificare le interrogazioni.

Pertanto, in questa sezione del DMP occorre indicare:

- quale interfaccia utente/portale/sito web verrà utilizzata per la pubblicazione dei dati
- chi avrà in carico la gestione dell'interfaccia utente
- qual è la localizzazione dell'interfaccia utente (URL o indirizzo IP)
- se l'interfaccia utente rispetta o meno i requisiti di accessibilità previsti dalla legge 9 gennaio 2004, n. 4 (*Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici*)
- se l'interfaccia utente rispetta o meno i requisiti di usabilità previsti nel Piano quinquennale per l'informatica nella Pubblica Amministrazione²⁰
- se è prevista o meno la pubblicazione nel portale nazionale <https://dati.beniculturali.it> (cfr. par. 7.5), di quali dati e con quali modalità e temporalità (conferimenti manuali o *harvesting*).

Inoltre, poiché ogni interfaccia utente ha generalmente un suo *backend* di gestione, nel DMP va inoltre indicato:

- quale struttura di *backend* verrà utilizzata per la pubblicazione dei dati e in che modo è integrata all'interfaccia utente (portale/sito web)
- chi avrà in carico la gestione del portale/sito web
- qual è la localizzazione fisica del portale/sito web (dove si trovano fisicamente i server che conservano i dati)
- qual è la localizzazione virtuale del portale/sito web (URL o indirizzo IP)
- in che modo il portale/sito web si relaziona con l'infrastruttura software del patrimonio culturale (cfr. par. 7.7.1)
- in caso di utilizzo dell'infrastruttura software del patrimonio culturale gestita dall'ICDP, qual è il modello di interazione adottato (modello integrato/modello federato²¹).

²⁰ <https://www.agid.gov.it/it/design-servizi/usabilita>

²¹ I due principali modi per relazionarsi con l'infrastruttura software del patrimonio culturale sono: a) *modello integrato*: gli enti conferiscono i propri dati all'infrastruttura, condividendone i servizi: le risorse digitali sono quindi "ospitate" nell'infrastruttura software e vengono memorizzate e conservate sui sistemi dell'infrastruttura, laddove il ciclo di vita della risorsa digitale viene gestito tramite i servizi dell'infrastruttura; b) *modello federato*: gli enti che hanno sistemi informativi in grado di esporre in modo stabile ed efficiente le risorse digitali mediante API standard, possono condividere con l'infrastruttura solo alcuni servizi, in base alle loro specifiche necessità; le risorse digitali risiedono nei sistemi di origine e sono "referenziate" nell'infrastruttura, mentre il ciclo di vita dei dati è gestito dall'ente nei propri sistemi.

7.7.3. Preservazione dei dati ⓘ

Nella gestione di un dataset, è fondamentale garantire la conservazione a lungo termine del dataset stesso. Il termine “*digital preservation*” indica le strategie e i metodi adottati per garantire la conservazione del dato nel corso del tempo²².

Sia i formati utilizzati per i dati e i metadati, sia i supporti sui quali il dataset risiede, sono soggetti a obsolescenza, pertanto è molto importante analizzare sin dall’inizio del progetto di digitalizzazione i rischi in cui è possibile incorrere e le tecniche che possono essere adottate per garantire la *digital preservation*.

In questa sezione sono descritti i possibili rischi per la preservazione a lungo termine dei dati, e in particolare:

- se e quando è prevista una futura obsolescenza delle tecnologie utilizzate (ad esempio, un determinato software non più aggiornato)
- se e quando è prevista una futura obsolescenza dei formati utilizzati per rappresentare i dati (ad esempio in seguito a sostituzione con nuovi formati)
- se sono previsti i rischi legati alla conservazione fisica del dato (ad esempio perdita di dati in seguito a incidente)
- quali altri rischi relativi alla preservazione dei dati sono previsti
- con quali modalità si prevede di far fronte ai rischi di cui sopra.

7.7.4. Procedure di *backup* ⓘ ⓘ

Oltre alla gestione dei rischi legati alla conservazione, è necessario individuare e implementare specifiche procedure di duplicazione (*backup*) per evitare possibili perdite di dati in caso di incidente. Tali procedure dipendono da come sono strutturati i sistemi di gestione dei dati utilizzati (sistemi *on-premise*, sistemi in *hosting* o in *housing*, sistemi in *cloud*).

In questo paragrafo vanno descritte le procedure di *backup* dei dati e dei metadati e le modalità di recupero previste in caso di incidente. In particolare, occorre indicare:

- quali sono le procedure di *backup* dei dati e metadati
- dove sono conservati i *backup*

²²Lee, K. H., Slattery, O., Lu, R., Tang, X., & McCrary, V. (2002). The state of the art and practice in digital preservation. *Journal of research of the National institute of standards and technology*, 107(1), 93.

- con quale frequenza vengono eseguiti i *backup*
- quali modalità di recupero dei dati sono previste in caso di incidente
- chi ha la responsabilità di eseguire i *backup*.

7.7.5. Certificazione per la conservazione a lungo termine ⓘ

Le eventuali attività di certificazione previste per la conservazione a lungo termine possono essere documentate e, in particolare, possono essere segnalate:

- quali attività di certificazione sono previste
- qual è l'ente responsabile della certificazione
- che tipo di controlli sono previsti
- chi ha la responsabilità di gestire la certificazione.

7.8. Sicurezza dei dati

La digitalizzazione dei dati semplifica l'accesso e la condivisione degli stessi, ma introduce anche una serie di rischi legati alla sicurezza informatica (o *cybersecurity*).²³ È infatti possibile che i dati condivisi in rete, se non adeguatamente protetti, subiscano furti, compromissioni o alterazioni da parte di soggetti malintenzionati.

L'Agenzia per l'Italia Digitale ha previsto una serie di misure atte a evitare possibili accessi non autorizzati o violazioni dei privilegi di accesso, valutando periodicamente le possibili vulnerabilità che potrebbero condurre a violazioni della sicurezza dei dati. In particolare, sono previsti tre diversi livelli di attuazione²⁴:

- **Minimo:** è quello al quale ogni Pubblica Amministrazione, indipendentemente dalla sua natura e dimensione, deve necessariamente essere o rendersi conforme.
- **Standard:** è il livello, superiore al livello minimo, che ogni amministrazione deve considerare come base di riferimento in termini di sicurezza e rappresenta la maggior parte delle realtà della PA italiana.
- **Avanzato:** deve essere adottato dalle organizzazioni maggiormente esposte a rischi (ad esempio per la criticità delle informazioni trattate o dei servizi erogati), ma anche visto come obiettivo di miglioramento da parte di tutte le altre organizzazioni.

Il MiC consiglia l'adozione delle misure di sicurezza standard, ma è in ogni caso indispensabile raggiungere almeno il livello minimo. Un approfondimento sul tema è fornito dalle misure di sicurezza ICT per le pubbliche amministrazioni indicate dall'AgID.²⁵

²³ Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, 80(5), 973-993.

²⁴ <https://www.agid.gov.it/it/sicurezza/misure-minime-sicurezza-ict>

²⁵ Agenzia per l'Italia Digitale (2016), *Misure minime di sicurezza ICT per le pubbliche amministrazioni* (https://www.agid.gov.it/sites/default/files/repository_files/documentazione/misure_minime_di_sicurezza_v.1.0.pdf)

7.8.1. Disposizioni per la sicurezza dei dati ⓘ

Nella sezione sono descritte le misure che verranno messe in atto per il rispetto delle misure minime di sicurezza ICT per le pubbliche amministrazioni, in base al livello AgID che l'istituto ha deciso di adottare.

Si consiglia inoltre di compilare la checklist per la verifica degli AgID Basic Security Controls – ABSC (si veda il par. 7.8.2) al fine di verificare il rispetto dei requisiti minimi di sicurezza. Gli aspetti che sono da considerare in questa sezione:

- se sono rispettati o meno i requisiti di sicurezza minimi ABSC
- se sono rispettati o meno requisiti di sicurezza di livello superiore
- quali sono le modalità di verifica periodica dei requisiti
- con che frequenza verranno eseguite queste verifiche
- chi ha la responsabilità di eseguire queste verifiche.

7.8.2. Standard per la sicurezza dei dati ⓘ

Per garantire la sicurezza dei dati, è opportuno adottare criteri standard di valutazione della sicurezza. Il MiC consiglia di adottare lo standard ABSC (AgID Basic Security Controls) definito dall'Agenzia per l'Italia digitale²⁶, che è compatibile con il framework FNSC (Framework Nazionale di Sicurezza Cibernetica²⁷) e con lo standard internazionale CCSC (CIS Critical Security Controls²⁸). Lo standard ABSC prevede la creazione di un inventario dei dispositivi autorizzati e non autorizzati, l'adozione di standard per la protezione dei dispositivi, una valutazione continua delle vulnerabilità per far fronte a nuovi possibili rischi, un uso appropriato dei privilegi di accesso e la protezione da *malware*. Sono inoltre previste altre due attività molto importanti, ovvero l'esecuzione di copie di *backup* (cfr. par. 7.7.5) e la protezione dei dati soggetti a vincoli di riservatezza (sicfr. Par. 7.8.4).

In questa sezione del DMP è opportuno riportare:

- se è stato adottato o meno lo standard ABSC
- se no, per quale motivo
- quali altri standard di sicurezza sono stati adottati
- in che modo è stato verificato il rispetto di tali standard.

²⁶ Questo standard è descritto nelle *Misure minime di sicurezza ICT per le pubbliche amministrazioni* citate in precedenza.

²⁷ <https://www.cybersecurityframework.it>

²⁸ <https://www.cisecurity.org/controls>

7.8.3. Rischi per la sicurezza dei dati ⓘ

Questa sezione delinea sinteticamente i possibili rischi previsti per la sicurezza dei dati, facendo riferimento alla specifica tipologia di dati raccolti nel progetto. In particolare, occorre indicare, se si prevedono:

- rischi di furto dei dati
- rischi di riuso improprio o illegale dei dati
- violazioni delle licenze adottate
- violazioni delle *policy* di accesso
- attacchi informatici all'infrastruttura
- attacchi informatici all'interfaccia utente
- qualunque altro rischio per la sicurezza dei dati non elencato sopra.

7.8.4. Sicurezza dei dati sensibili ⓘ

Nel caso di pubblicazione di dati soggetti al Regolamento sulla protezione dei dati personali, occorre prevedere una regolamentazione degli accessi che potrà fare capo a un sistema di registrazione online dell'utenza mediante SPID, in modo da riconoscere con certezza l'identità del richiedente e informarlo adeguatamente in merito alle regole deontologiche per il trattamento dei dati personali e alle responsabilità derivanti da un utilizzo illegittimo degli stessi. In alcuni casi può essere opportuno prevedere forme di anonimizzazione o pseudonimizzazione dei dati.

In questa sezione del DMP occorre quindi indicare:

- quali misure vengono adottate nel progetto per garantire la sicurezza dei dati sensibili
- quali forme di controllo degli accessi sono previste
- quali livelli di accesso sono previsti
- quali modalità di autenticazione sono previste
- se è stata adottata l'autenticazione tramite SPID
- se non è stata adottata, indicare per quale motivo
- se sono previste forme di registrazione *online* degli utenti
- se sì, quali modalità di registrazione sono previste e con quali modalità verranno conservati i dati
- se è stata eseguita una anonimizzazione o pseudonimizzazione dei dati, e se sì di quali dati e con quali modalità.

8. Strumenti per la compilazione del DMP ⓘ

[in corso di redazione]

Normativa di riferimento

Questa sezione elenca la normativa di riferimento rilevante per la redazione del DMP.

Riferimenti normativi italiani

- Decreto legislativo 30 giugno 2003, n. 196 – Codice in materia di protezione dei dati personali
- Decreto legislativo 7 marzo 2005, n. 82 – Codice dell'amministrazione digitale (in breve CAD)
- Decreto legislativo 24 gennaio 2006, n.36 – Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico
- Decreto legislativo 27 gennaio 2010, n. 32 – Attuazione della direttiva 2007/2/CE, che istituisce un'infrastruttura per l'informazione territoriale nella Comunità europea (INSPIRE)
- 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 (Decreto trasparenza)
- Decreto legislativo 18 maggio 2015, n.102 – Attuazione della direttiva 2013/37/UE relativa al riutilizzo di documenti nel settore pubblico
- Decreto della Presidenza del Consiglio dei Ministri 10 novembre 2011 – Regole tecniche per la definizione del contenuto del Repertorio nazionale dei dati territoriali, nonché delle modalità di prima costituzione e di aggiornamento dello stesso
- Legge 1° ottobre 2020 n. 133 – Ratifica ed esecuzione della Convenzione quadro del Consiglio d'Europa sul valore del patrimonio culturale per la società, fatta a Faro il 27 ottobre 2005

Riferimenti normativi europei

- Regolamento (CE) 2008/1205 del 3 dicembre 2008 recante attuazione della direttiva 2007/2/CE del Parlamento europeo e del Consiglio per quanto riguarda i metadati

- Regolamento (UE) 2010/1089 del 23 novembre 2010 recante attuazione della direttiva 2007/2/CE del Parlamento europeo e del Consiglio per quanto riguarda l'interoperabilità dei set di dati territoriali e dei servizi di dati territoriali
- Regolamento (UE) 2016/679 del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali (in breve GDPR)
- Direttiva (UE) 2019/1024 del 20 giugno 2019 relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico
- Decisione (UE) 2019/1372 del 19 agosto 2019 recante attuazione della direttiva 2007/2/CE del Parlamento europeo e del Consiglio per quanto riguarda il monitoraggio e la comunicazione
- Comunicazione della Commissione 2014/C 240/01 del 24 luglio 2014 – Orientamenti sulle licenze standard raccomandate, i dataset e la tariffazione del riutilizzo dei documenti
- Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni COM(2020) del 19 febbraio 2020 – Una strategia europea per i dati
- Conclusioni del Consiglio del 21 maggio 2014 relative al patrimonio culturale come risorsa strategica per un'Europa sostenibile
- Raccomandazione (UE) 2021/1970 della Commissione del 10 novembre 2021 relativa a uno spazio comune europeo di dati per il patrimonio culturale

Linee guida nazionali

- Agenzia per l'Italia Digitale, *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico*, <https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/>
- Agenzia per l'Italia Digitale, *Misure minime di sicurezza ICT per le pubbliche amministrazioni*, <https://www.agid.gov.it/it/sicurezza/misure-minime-sicurezza-ict>
- Agenzia per l'Italia Digitale, *Piano triennale per l'informatica nella Pubblica Amministrazione*, <https://www.agid.gov.it/it/agenzia/piano-triennale>

Allegato: FAQ per la pubblicazione dei dati aperti

Le FAQ che seguono sono state elaborate a partire dalle *Linee guida per valorizzazione del patrimonio informativo pubblico*, a cura dell'AgID e del Dipartimento per la trasformazione digitale della Presidenza del Consiglio dei Ministri²⁹ (nel seguito anche "Linee guida AgID"). Tali Linee guida hanno l'obiettivo di supportare gli uffici nel processo di valorizzazione del proprio patrimonio informativo pubblico, nel rispetto degli obiettivi indicati nell'articolo 52 del D.lgs 7 marzo 2005, n. 82 - Codice dell'Amministrazione Digitale (CAD).

Le linee guida propongono a tal fine un modello e un'architettura di riferimento per l'informazione del settore pubblico, individuando:

- standard di base, formati, vocabolari/ontologie per dati di specifici domini
- profili di metadati descrittivi nazionali
- aspetti organizzativi necessari per individuare i ruoli e le figure professionali
- fasi dei processi per la gestione e pubblicazione di dati di qualità.

Inoltre, il documento mira a fornire supporto: nella scelta della licenza per i dati di tipo aperto, nell'analisi di eventuali aspetti di costo dei dati e nella loro pubblicazione nei portali per una maggiore standardizzazione di questo processo.

Le linee guida AgID hanno una duplice valenza tecnica e organizzativa e si rivolgono sia a figure professionali delle amministrazioni in possesso di competenze tecnico-informatiche (ad esempio, direttori dei sistemi informativi, responsabili siti web, funzionari e consulenti tecnici), sia a figure professionali individuabili in quelle aree più amministrative preposte all'organizzazione dei dati (ad esempio, responsabili di basi di dati specifiche, responsabili amministrativi, esperti di dominio).

²⁹ <https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/index.html>.

Cosa sono i dati aperti (open data)?

Il Codice dell'Amministrazione digitale (decreto legislativo 7 marzo 2005, n. 82) all'art. 1, definisce aperti i dati:

- disponibili secondo i termini di una licenza o di una previsione normativa che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato;
- accessibili attraverso le tecnologie dell'informazione e della comunicazione [...] in formati aperti[...];
- adatti all'utilizzo automatico da parte di programmi per elaboratori;
- provvisti dei relativi metadati;
- resi disponibili gratuitamente oppure resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

Con dato aperto si intende dunque un dato che risponde ai seguenti principi di base:

- Disponibile (requisito giuridico) secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato;
- Accessibile (requisito tecnologico) attraverso le tecnologie dell'informazione e della comunicazione, in formato aperto e con i relativi metadati;
- Gratuito (requisito economico): disponibili gratuitamente oppure disponibili ai costi marginali sostenuti per la loro riproduzione, messa a disposizione e divulgazione.

I dati aperti del MiC aderiscono ai principi FAIR.

Cosa sono i principi FAIR?

Nel 2014 sono stati elaborati alcuni principi per la condivisione dei dati scientifici, denominati FAIR (acronimo derivante dai termini *Findable, Accessible, Interoperable, Re-Usable*), per ottimizzare la riutilizzabilità dei dati e dei risultati della ricerca. Tali principi, ormai riconosciuti a livello internazionale, descrivono le caratteristiche che le risorse digitali debbono avere per essere usate e riutilizzate a fini scientifici, educativi e divulgativi, sia dalle persone sia dalle macchine che adottano processi automatizzati.

Rintracciabilità (*Findability*)

I dati e i relativi metadati raccolti nel progetto devono essere facilmente rintracciabili sia da parte di utenti umani sia da parte di strumenti informatici. Nel DMP va specificato in che modo il progetto soddisfa i requisiti di rintracciabilità, e in particolare tramite:

- **Identificatori univoci e persistenti:** è opportuno valutare che gli identificatori utilizzati identifichino in maniera univoca le risorse, che non siano soggetti a cambiamento nel tempo, e che i metadati siano collegati in maniera esplicita agli identificatori delle risorse.
- **Metadati sufficientemente descrittivi:** valutare se i metadati descrivono in modo adeguato le risorse. Inoltre, i metadati devono essere collegati in maniera esplicita agli identificatori univoci.
- **Indicizzazione dei metadati:** è opportuno verificare che i metadati siano indicizzati in risorse disponibili pubblicamente (es. messe a disposizione dal MiC), in modo che sia semplice reperire le risorse.

Accessibilità (*Accessibility*)

I dati e i relativi metadati raccolti nel progetto devono essere facilmente accessibili tramite la pubblicazione in repository aperti e disponibili pubblicamente. Nel DMP va specificato in che modo il progetto soddisfa i requisiti di accessibilità, e in particolare tramite:

- **Pubblicazione in repository aperti (*open access*):** è opportuno verificare che i dati siano pubblicati in repository aperti, in modo che sia semplice per qualunque utente reperire le risorse, e che sia consentito il riuso dei dati.
- **Protocollo di accesso:** è opportuno verificare che il dataset venga reso accessibile tramite un protocollo aperto (es. HTTP, FTP).
- **Persistenza dei metadati:** è opportuno verificare che venga garantita la persistenza dei metadati indipendentemente dall'accessibilità del dataset e indicare se il requisito è soddisfatto.
- **Sistemi di autorizzazione e autenticazione:** qualora fossero presenti restrizioni sull'accesso ai dati, è opportuno verificare che siano supportati meccanismi di autorizzazione e autenticazione per l'accesso.

Interoperabilità (*Interoperability*)

I dati e i relativi metadati raccolti nel progetto devono garantire l'interoperabilità, in modo da facilitare il più possibile l'integrazione con altri sistemi e dataset. Nel DMP va specificato in che modo il progetto soddisfa i requisiti di interoperabilità, e in particolare tramite:

- Adozione di standard open data: è opportuno valutare se gli standard e i formati adottati nel progetto garantiscono l'interoperabilità del dataset.
- Adozione di vocabolari compatibili con i principi FAIR: è opportuno valutare se i vocabolari adottati nel progetto garantiscono l'interoperabilità del dataset.
- Integrazione con altri dataset: è opportuno valutare se è prevista un'integrazione o *linking* con altri dataset, e con quali modalità.

Riutilizzabilità (*Reusability*)

I dati e i relativi metadati raccolti nel progetto devono garantire la riutilizzabilità tramite l'adozione di standard, licenze aperte e indicazioni sull'origine dei dati. In questo campo va specificato in che modo il progetto soddisfa i requisiti di riutilizzabilità, e in particolare tramite:

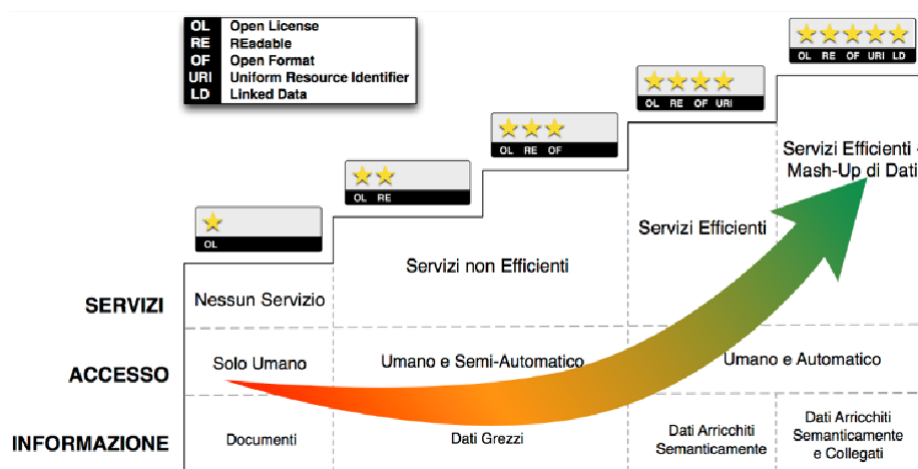
- Adozione di licenze aperte: è opportuno valutare se le licenze adottate rispettano la definizione di licenza aperta (a questo proposito, si vedano le *Linee guida per l'acquisizione, la circolazione e il riuso delle riproduzioni dei beni culturali in ambiente digitale*).
- Metadati di *provenance*: è opportuno valutare se i metadati descrivono in maniera adeguata l'origine dei dati e i processi di produzione, raccolta e trasformazione di cui i dati sono stati oggetto.
- Standard per i metadati: è opportuno valutare se le metodologie di metadattazione adottate nel progetto rispettano gli standard di settore.

Cosa significa formato aperto (#FormatoAperto)?

Con formato dei dati di tipo aperto si intende un formato reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi.

Cosa sono i livelli degli open data?

Le Linee guida AgID, che riprendono la definizione di Open Data del W3C³⁰, prevedono cinque possibili livelli di open data, come di seguito mostrato:



La pubblicazione di dati di livello 1 e 2 non è più consentita; il MiC pertanto pubblicherà i dati in formato aperto almeno di livello 3* mirando, nel tempo, a pubblicare solo dati di livello a 4* e 5* (*Linked Open Data*).

Cosa sono i linked (open) data?

I *linked data* (trad.: dati collegati [ad altri dati]) sono una modalità di pubblicazione di dati strutturati basata su tecnologie e standard aperti del web come HTTP, RDF (*Resource Description Framework*) e URI (*Uniform Resource Identifier*). Se i *linked data* collegano dati aperti si parla di *linked open data* (LOD). I *linked data* sono una delle tecnologie alla base del cosiddetto *Semantic Web* (una sorta di spazio globale dei dati interconnessi tra loro con relazioni semanticamente qualificate) in cui i dati, strutturati e collegati tra loro, costruiscono un reticolo informativo sempre più ampio che i software riescono a leggere e interpretare direttamente estraendo informazioni attraverso interrogazioni di tipo semantico. I dati e le relazioni tra essi sono descritti semanticamente tramite metadati e ontologie. Nel collegare (o referenziare) si usano dunque relazioni ("link") che hanno un preciso significato e spiegano il tipo di legame che intercorre tra le due entità coinvolte nel collegamento. I *linked open data* sono quindi un metodo elegante ed efficace per risolvere problemi di identità e provenienza, semantica, integrazione e interoperabilità.

³⁰ <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html#x5-star-linked-open-data>

Già nel 2012 la Commissione di Coordinamento SPC ha emanato le Linee guida sull'interoperabilità semantica attraverso i *linked open data* e la metodologia ivi proposta risulta essere ancora valida anche se la filiera di lavorazione dei LOD è un percorso che può essere complesso da intraprendere e che richiede competenze tecniche specifiche. Tuttavia, l'intenzione del MiC è governare una transizione graduale verso la produzione di LOD che sarà possibile se trainata dagli Istituti centrali e dalle Direzioni generali.

Esistono leggi che impongono di pubblicare dati in formato aperto?

Il principio dell'*open data by default* è stato introdotto nel 2012 con la modifica dell'articolo 52 del Codice dell'Amministrazione Digitale, per cui "i dati e i documenti che [le PA] pubblicano, con qualsiasi modalità, senza l'espressa adozione di una licenza [...] si intendono rilasciati come dati di tipo aperto [...] ad eccezione dei casi in cui la pubblicazione riguardi dati personali [...]". Altri importanti cambiamenti normativi sono intervenuti soprattutto per quel che riguarda il recepimento della nuova direttiva Europea 2019/1024, che abroga la vigente direttiva 2003/98/CE apportando significative novità in tema di riutilizzo dei dati aperti della PA: essa accresce la rilevanza economica del riutilizzo degli open data andando ad estenderne il campo di applicazione alle attività di interesse economico generale, ai "dati dinamici" e "di elevato valore" nonché ai dati prodotti nell'ambito della ricerca scientifica.

La direttiva 2019/1024 sul riutilizzo dei dati prodotti dalla pubblica amministrazione (Public Sector Information) è la terza direttiva PSI in ordine di tempo: la direttiva PSI del 2013, rispetto alla prima del 2003, ha incluso per la prima volta nel proprio ambito di applicazione i dati detenuti musei, archivi e biblioteche, ma ha finito per rimanere in buona parte priva di effetti su questi istituti per la mancata emanazione del decreto ministeriale che avrebbe dovuto definire i criteri di tariffazione legati al riutilizzo dei dati detenuti dagli istituti pubblici di tutela.

Tutti i dati devono essere pubblicati in formato aperto?

Tutti i dati detenuti dalle pubbliche amministrazioni devono essere pubblicati in formato aperto, con le seguenti esclusioni:

- dati a conoscibilità limitata come i dati coperti da segreto di stato o le opere d'ingegno coperte dal diritto d'autore;
- i dati personali, per i quali trovano applicazione le norme del «Codice in materia di protezione dei dati personali» (i.e., D.lgs n. 196/2003 e Linee guida in materia di trattamento di dati personali e s.m.i.). In questo caso, si ponga anche attenzione a non esporre quasi-identificatori (e.g., data di nascita, domicilio, residenza, sesso, razza, etnia, composizione nucleo familiare, status giuridico, ecc.) che possono facilmente re-identificare i soggetti che si intende invece tutelare o che hanno una tutela speciale perché appartenenti a fasce protette (e.g., testimoni giudiziari, profughi, rifugiati, pentiti, ecc.). In ogni caso, AgID raccomanda di verificare gli artt. 3 e 4 del D. Lgs. 36/2006 per una visione approfondita circa le esclusioni e le norme di salvaguardia.

Per quanto riguarda le riproduzioni digitali dei beni culturali un limite alla pubblicazione in formato aperto è rappresentato dalle prescrizioni dell'art. 107 e 108 del Codice dei beni culturali, dove si prevede la corresponsione di un canone per i riusi commerciali; nessuna limitazione è invece posta dal Codice alla pubblicazione in formato aperto dei dati descrittivi del patrimonio culturale. Per un approfondimento sul tema si rimanda alle *Linee guida per l'acquisizione, la circolazione e il riuso delle riproduzioni di beni culturali in ambiente digitale*.

AgID fornisce una breve "check list", utile per verificare se tutti gli aspetti giuridici sono stati valutati dal responsabile della banca dati. La check list è formata da una serie di domande, per ciascun aspetto, a cui rispondere con Sì o No.

AMBITO	DOMANDA DI CONTROLLO	SI/NO
<i>Privacy</i>	I dati sono liberi da ogni informazione personale che possa identificare in modo diretto l'individuo (nome, cognome, indirizzo, codice fiscale, patente, telefono, email, foto, descrizione fisica, ecc.)? In caso negativo queste informazioni sono autorizzate per legge?	
	I dati sono liberi da ogni informazione indiretta che possa identificare l'individuo (caratteristiche personali che possono identificare facilmente il soggetto)? In caso negativo queste informazioni sono autorizzate per legge?	
	I dati sono liberi da ogni informazione sensibile riconducibile all'individuo? In caso negativo queste informazioni sono autorizzate per legge?	
	I dati sono liberi da ogni informazione relativa al soggetto che incrociata con dati comunemente reperibili nel web (e.g. google maps, linked data, ecc.) possa identificare l'individuo? In caso negativo queste informazioni sono autorizzate per legge?	
	I dati sono liberi da ogni riferimento a profughi, protetti di giustizia, vittime di violenze o in ogni caso categorie protette?	
	Hai considerato il rischio di de-anonimizzazione del tuo dataset prima di pubblicarlo?	
	Esponi dei servizi di ricerca tali da poter filtrare i dati in modo da ottenere un solo record geolocalizzato, che sia facilmente riconducibile ad una persona fisica?	
Proprietà intellettuale della sorgente	Il dataset è stato creato da uno o più dipendenti della tua pubblica amministrazione nell'ambito della loro attività lavorativa? I singoli elementi del dataset suscettibili di autonoma protezione (es., immagini, fotografie, testi in qualche modo creativi) sono stati a loro volta prodotti da uno o più dipendenti della tua pubblica amministrazione nell'ambito della loro attività lavorativa?	
	L'amministrazione è proprietaria 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 della tua amministrazione hai un accordo o una licenza che ti autorizzi a pubblicarli?	

Licenza di rilascio	Stai rilasciando i dati di cui possiedi la proprietà accompagnati da una licenza?	
	Hai incluso anche la clausola di salvaguardia «Questo dataset contiene informazioni indirettamente riferibili a persone fisiche. In ogni caso, i dati non possono essere utilizzati al fine di identificare nuovamente gli interessati.»?	
Limiti alla pubblicazione	Hai verificato che non vi siano impedimenti di legge o contrattuali che per la pubblicazione dei dati?	
Segretezza	Hai verificato se non vi siano motivi di ordine pubblico o di sicurezza nazionale che ti impediscono la pubblicazione dei dati?	
	Hai verificato se non vi siano motivi legati al segreto d'ufficio che impediscono la pubblicazione dei dati?	
	Hai verificato se non vi siano motivi legati al segreto di stato che impediscono la pubblicazione dei dati?	
Indicazioni temporali	I dati sono soggetti per legge a restrizioni temporali di pubblicazione?	
	I dati sono aggiornati frequentemente in modo da sanare eventuali informazioni lesive di persone o organizzazioni?	
	I dati hanno dei divieti di legge o giurisprudenziali che impediscono la loro indicizzazione da parte di motori di ricerca?	
Trasparenza	I dati rientrano nella lista dell'allegato A del d.lgs. 33/2013? Se sì come sono stati trattati dal responsabile della trasparenza nella sezione "Amministrazione trasparente"?	

E se i dati contengono riferimenti espliciti a persone (dato personale)?

In questo caso i dati non vanno pubblicati in formato aperto, a meno che non sia possibile procedere all'anonimizzazione del dato. I dati possono essere considerati anonimi quando le persone non sono più identificabili. Infatti, esistono molte altre informazioni che consentono a un individuo di essere collegato ai suoi dati personali e che ne consentono pertanto la reidentificazione. Il GDPR, però, non prescrive alcuna tecnica particolare per l'anonimizzazione; spetta quindi ai singoli responsabili del trattamento

garantire che qualunque processo di anonimizzazione scelto sia sufficientemente solido.

Che vantaggi si traggono dalla pubblicazione dei dati aperti?

La valorizzazione del patrimonio informativo pubblico è un obiettivo strategico per la pubblica amministrazione, soprattutto per affrontare efficacemente le nuove sfide dell'economia dei dati (*data economy*), supportare la costruzione del mercato unico europeo per i dati definito dalla Strategia europea in materia di dati³¹, garantire la creazione di servizi digitali a valore aggiunto per cittadini, imprese e, in generale, tutti i portatori di interesse e fornire al decisore politico strumenti *data-driven* da utilizzare nei processi decisionali.

A tal fine, il Piano triennale per l'informatica nella Pubblica Amministrazione ridefinisce una nuova *data governance* coerente con la Strategia europea e con il quadro delineato dalla nuova Direttiva europea sull'apertura dei dati e il riutilizzo dell'informazione del settore pubblico.

Il principio generale della direttiva è quello di favorire al massimo il riutilizzo dei dati della pubblica amministrazione, a eccezione dei dati esclusi dal diritto di accesso ai sensi del diritto nazionale e in conformità alla normativa sulla protezione dei dati. Questo principio muove dalla convinzione che il libero riutilizzo dei dati, anche per fini commerciali, è un potente moltiplicatore di ricchezza e un asset strategico per lo sviluppo sociale, culturale ed economico dei Paesi membri in una fase di forte crescita dei settori che si occupano dell'elaborazione di dati grezzi in materiale per lo sviluppo di nuove app e servizi che possono essere erogati da soggetti pubblici e privati: maggiore è infatti la qualità e quantità degli Open Data messi a disposizione dalle pubbliche amministrazioni, e maggiori saranno le probabilità che i dati verranno utilizzati al fine di creare servizi innovativi capaci di divenire fattori di benessere per la società.

Per tali ragioni già la direttiva del 2013 prescriveva l'obbligo, e non più solamente la facoltà, per le amministrazioni, di rendere riutilizzabili per fini commerciali o non commerciali i dati in loro possesso, ove possibile per via elettronica e in formati aperti, leggibili meccanicamente, accessibili, reperibili e riutilizzabili, insieme ai rispettivi metadati.

³¹ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_it

Si possono fare pagare i dati?

La condivisione dei dati tra pubbliche amministrazioni per finalità istituzionali (art. 50 del CAD), avviene esclusivamente a titolo gratuito. Anche nel caso della pubblicazione di *open data*, AgID suggerisce di renderli disponibili esclusivamente a titolo gratuito. Tuttavia, è prevista la possibilità di richiedere per il riutilizzo dei dati un corrispettivo specifico, limitato ai costi sostenuti effettivamente per la riproduzione, messa a disposizione e divulgazione dei dati. In tali casi, come previsto dall'art. 7 del D.Lgs 24 gennaio 2006, n. 36, AgID determina le tariffe standard da applicare, pubblicandole sul proprio sito istituzionale. Nel pieno rispetto dei principi di trasparenza e verificabilità, tali tariffe sono determinate sulla base del "Metodo dei costi marginali" esplicitato nella Comunicazione della Commissione 2014/C - 240/01 contenente, tra gli altri, gli orientamenti sulla tariffazione. In linea con quanto previsto dalla direttiva comunitaria, il citato articolo 7 del D.Lgs. 36/2006 prevede inoltre casi specifici per i quali è possibile determinare tariffe superiori ai costi marginali in deroga al principio generale di rendere disponibili i dati gratuitamente o a costi marginali, ovvero:

- alle biblioteche, comprese quelle universitarie, di musei e archivi;
- alle amministrazioni e agli organismi di diritto pubblico che devono generare utili per coprire una parte sostanziale dei costi inerenti allo svolgimento dei propri compiti di servizio pubblico;
- ai casi eccezionali relativi a documenti per i quali le pubbliche amministrazioni e gli organismi di diritto pubblico sono tenuti a generare utili sufficienti per coprire una parte sostanziale dei costi di raccolta, produzione, riproduzione e diffusione.

Alla data di elaborazione del presente documento sono in corso di redazione da parte di AgID i criteri per la determinazione di tali tariffe.

Per quanto riguarda invece i criteri per la tariffazione delle riproduzioni dei beni culturali, si rimanda a quanto previsto nelle *Linee guida per l'acquisizione, la circolazione e il riuso delle riproduzioni digitali dei beni culturali in ambiente digitale*.

Con che licenza si devono pubblicare i dati aperti (#Licenza)?

Per licenza d'uso si intende il contratto, o altro strumento negoziale, redatto ove possibile in forma elettronica, nel quale sono definite le modalità di riutilizzo dei documenti delle pubbliche amministrazioni o degli organismi di diritto pubblico.

L'informazione sul tipo di licenza è un metadato indispensabile per determinare come poter riutilizzare il dataset (ovvero l'insieme organico dei dati resi disponibili). Deve pertanto essere sempre specificata indicando, il nome, la versione e fornendo il riferimento al testo della licenza.

Nel contesto dei dati aperti, considerando la definizione *Open Data* fornita dal CAD e dall'Open Knowledge Foundation (OKFN), per cui un dato è aperto se è "liberamente usabile, riutilizzabile e ridistribuibile da chiunque per qualsiasi scopo, soggetto al massimo alla richiesta di attribuzione e condivisione allo stesso modo", non tutte le licenze d'uso sono compatibili con i principi dei dati aperti. Nella figura che segue le licenze vengono classificate secondo tale criterio:

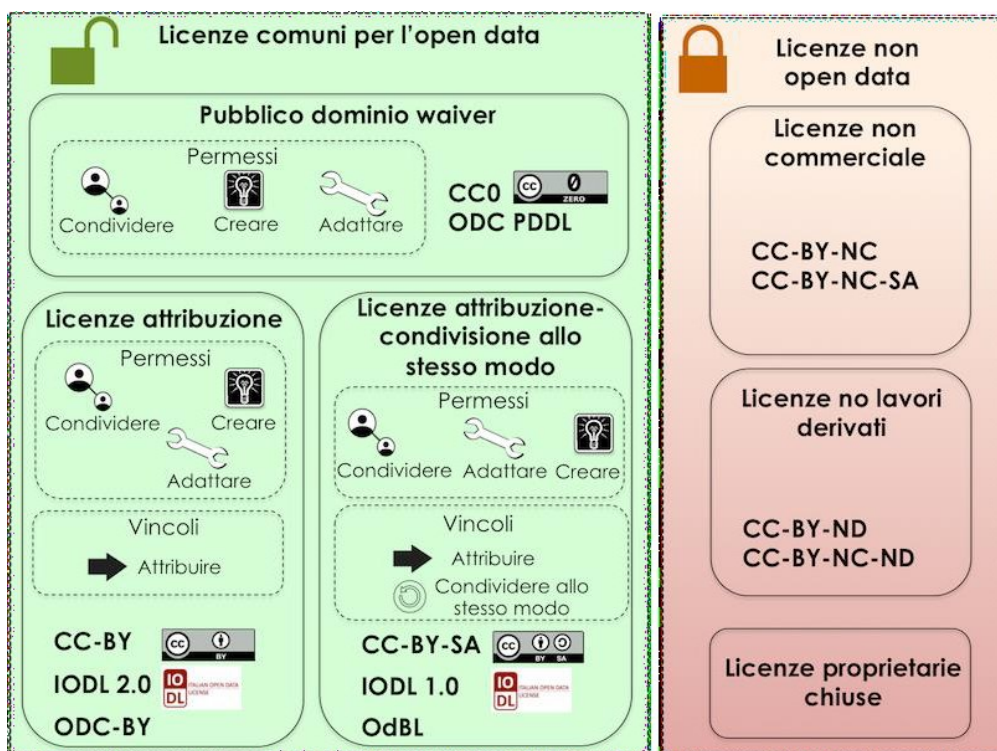


Figura 5 - Lo schema è tratto dalla figura disponibile al seguente link: <https://docs.italia.it/italia/daf/lq-patrimonio-pubblico/it/stabile/licenzecosti.html#id5>

Tutte le licenze che non consentono lavori derivati, anche per finalità commerciali, i.e., licenze che riportano chiaramente clausole Non Commercial - NCE/o Non Derivative - NDE/o ogni altra clausola che limita la possibilità di riutilizzo e ridistribuzione dei dati, non possono essere ritenute valide per identificare dataset aperti.

Le licenze più usate per gli open data appartengono a tre categorie principali:

1. il pubblico dominio o “*waiver*” dove il dichiarante “apertamente, pienamente, permanentemente, irrevocabilmente e incondizionatamente rinuncia, abbandona e cede ogni proprio diritto d’autore e connesso, ogni relativa pretesa, rivendicazione, causa e azione, sia al momento nota o ignota (includendo espressamente le pretese presenti come quelle future) relativa all’opera”. Rientrano in questa categoria la CC0 della famiglia delle licenze internazionali Creative Commons e la Open Data Commons – Public Domain Dedication License (ODC-PDDL) per i dataset/database;
2. le licenze per l’*open data* con richiesta di attribuzione, che consentono di condividere, adattare e creare anche per finalità commerciali con il solo vincolo di attribuire la paternità del dataset. Rientrano in questa categoria la licenza CC-BY della famiglia Creative Commons, la IODL (Italian Open Data License) nella sua versione 2.0 e la Open Data Commons Attribution License (ODC-BY) per dataset/database.
3. le licenze per l’*open data* con richiesta di attribuzione e condivisione allo stesso modo, che consentono di condividere, adattare e creare anche per finalità commerciali nel rispetto però di due vincoli: a) attribuire la paternità del dataset; b) distribuire eventuali lavori derivati con la stessa licenza che governa il lavoro originale. Rientrano in questa categoria la licenza CC-BY-SA della famiglia Creative Commons la IODL nella sua versione 1.0 la Open Data Commons Open Database License (ODbL) utilizzata dal progetto OpenStreetMap (OSM).

In relazione a quanto sopra riportato, tenuto conto del contesto normativo di riferimento, si ritiene opportuno fare riferimento ad una licenza unica aperta per tutto il MiC, che garantisca libertà di riutilizzo, che sia internazionalmente riconosciuta e che consenta di attribuire la paternità dei dataset (attribuire la fonte). Pertanto, si suggerisce l’adozione generalizzata della licenza CC-BY nella sua versione 4.0, presupponendo altresì l’attribuzione automatica di tale licenza nel caso di applicazione del principio “Open Data by default”, espresso nelle disposizioni contenute nell’articolo 52 del CAD. Per le

immagini dei beni culturali, si rimanda a quanto previsto nelle *Linee guida per l'acquisizione, la circolazione e il riuso delle riproduzioni di beni culturali in ambiente digitale*.

Per finalità particolari, ad esempio per il conferimento dei dati a portali di valorizzazione del patrimonio culturale (cfr. Europea) o progetti collaborativi di divulgazione del sapere (cfr. Wikidata), se richiesto dall'adesione al portale o al progetto, il MiC può valutare l'opportunità di rilasciare alcuni dataset con le licenze richieste dai suddetti progetti e portali.

AgiD raccomanda inoltre di gestire l'attribuzione della fonte indicando il nome dell'organizzazione unitamente all'URL della pagina Web dove si trovano i dataset/contenuti da licenziare. Nell'applicazione della licenza si ricorda che non si può disporre/attribuire diritti più ampi rispetto alla licenza di partenza (e.g., non si può attribuire un pubblico dominio - o *waiver* - a un dataset ottenuto da una fonte a cui è associata una licenza che richiede attribuzione).




Infine, le amministrazioni possono prevedere casi di applicazione di licenze che limitino il riutilizzo dei dati se e solo se ciò si renda necessario per il rispetto di altre normative (e.g., norme in materia di protezione dei dati personali) e comunque motivando opportunamente la scelta.

A completamento dell'argomento, si evidenzia l'opportunità di verificare gli aspetti relativi a:

- titolarità dei dati secondo la competenza amministrativa
- elaborazione di un'opera derivata, con il conseguente onere di citazione della fonte originale del dataset e di specifica attribuzione all'opera derivata
- finalità per i quali i dati sono stati creati che eventualmente non consentono di renderli automaticamente disponibili in open data
- responsabilità del titolare rispetto al riutilizzo dei dati da parte di terzi e, nel caso, specificare una nota legale, che integra e accompagna la licenza.

Un'indicazione di compatibilità tra le licenze *Open Data* è riportata nella tabella seguente:

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

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

Cosa sono le licenze Creative Commons (CC)

Le licenze più note a livello internazionale sono le Creative Commons (CC)³², proposte nel 2002 da Lawrence Lessig, d'uso ormai sempre più comune nell'editoria, nel mercato audiovisivo e nelle pratiche di digitalizzazione delle collezioni museali in tutto il mondo; tali licenze favoriscono una gestione più flessibile e intuitiva dei diritti d'autore gravanti sulle opere rilasciate in rete mediante il ricorso a loghi internazionalmente riconosciuti e a metadati *machine-readable* in grado di rendere immediatamente comprensibili all'utente i termini di utilizzo dell'opera. Lo strumento della licenza ha dunque il pregio di permettere all'autore dell'opera, o comunque al titolare dei diritti di sfruttamento economici, una gestione più agile ed equilibrata dei propri diritti favorendo al tempo stesso un uso più responsabile e consapevole delle risorse digitali da parte del pubblico. Le licenze CC si basano sul concetto di "some rights reserved" (alcuni diritti riservati) in opposizione alla formula tradizionale "all rights reserved" (tutti i diritti riservati).

Le licenze CC sono complessivamente sei e derivano dalla combinazione dei seguenti quattro attributi:

- **Attribuzione/Attribution (BY):** l'utente è tenuto ad attribuire la paternità dell'opera nel modo indicato dall'autore stesso;
- **Non opere derivate/No Derivatives (ND):** l'opera non può essere alterata o modificata dall'utente in nessun modo, né utilizzata per crearne una simile. È alternativa alla SA;
- **Non commerciale/Non Commercial (NC):** l'opera non può essere sfruttata dall'utente per fini commerciali;

³² <https://creativecommons.it/chapterIT/>

- Condividi allo stesso modo/*Share Alike* (SA): l'opera può essere modificata e può circolare solo per il tramite di una licenza equivalente a quella originaria. È alternativa alla ND.

Oltre alle sei licenze autoriali Creative Commons mette a disposizione altri due strumenti specificatamente riservati alle opere in pubblico dominio: l'etichetta PDM (*Public Domain Mark*) e il dispositivo CC0. PDM è propriamente un'etichetta, non una licenza, concepita per comunicare che l'opera risulta priva di restrizioni sul piano del diritto d'autore note a livello internazionale. Il dispositivo CC0 è invece uno strumento, dotato di valore legale (a differenza di PDM), che permette all'autore di rinunciare a ogni diritto sulle opere prodotte, compreso quello di attribuzione espressa (BY). In questo modo l'opera entra nel pubblico dominio non già in seguito alla scadenza dei termini di protezione, bensì per scelta volontaria del suo autore.

Come definire una lista di priorità per pubblicare i dati in formato aperto?

La strategia nazionale per gli *open data* delineata a partire dal 2017 nel "Piano Triennale per l'informatica nella PA" suggerisce un percorso che passa attraverso varie fasi operative:

1. individuazione di basi di dati altamente affidabili ed essenziali per un elevato numero di procedimenti amministrativi (altrimenti dette basi di dati di interesse nazionale o *base register* secondo la terminologia prevista nell'ambito dell'European Interoperability Framework),
2. apertura, in *open data*, della gran parte dei dati prodotti dalle amministrazioni, nel rispetto degli ambiti di applicazione previsti dalle norme,
3. definizione di vocabolari controllati e modelli dati, da rendere disponibili in un apposito registro consultabile da chiunque,
4. messa a disposizione di uno spazio dei dati che offre, tra gli altri, un servizio di Open Data as a Service (ODasS) certificati.

Nel contesto dei dati aperti, la strategia complessiva a livello nazionale include inoltre la definizione di un "Paniere dinamico di dataset" (inserito nel piano triennale e aggiornabile di anno in anno) attraverso il quale è individuato un insieme di basi di dati, sia regionali, sia nazionali, che le amministrazioni intendono rendere disponibili in open data. All'interno del paniere si considerano altresì richieste specifiche di dati da aprire provenienti

da iniziative ufficiali con la cittadinanza (e.g., Open Government Partnership Forum).

Tali elementi costituiscono anche la base di riferimento per diverse azioni di monitoraggio che devono essere intraprese per dar seguito sia agli impegni assunti nell'ambito del piano triennale, sia alle disposizioni dell'articolo 52 del Codice dell'Amministrazione Digitale e della suddetta Direttiva PSI 2.0. Il MiC segnala annualmente ad AgID quali basi di dati nazionali metterà a disposizione in *linked open data*, tra quelle detenute dagli Istituti centrali e dalle Direzioni generali.

Cosa sono le ontologie e a cosa servono?

In informatica, un'ontologia è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse. Il termine ontologia formale è entrato in uso nel campo dell'intelligenza artificiale e della rappresentazione della conoscenza, per descrivere il modo in cui diversi schemi vengono combinati in una struttura dati contenente tutte le entità rilevanti e le loro relazioni in un dominio. I software usano le ontologie per vari scopi, tra cui il ragionamento induttivo, la classificazione, etc.

AgID raccomanda di modellare i dati sulla base dei vocabolari e ontologie di OntoPiA in larga parte allineati (collegati) a standard aperti del Web e disponibili in formati aperti standard sulla piattaforma <https://github.com/italia/>.

Gli uffici del MiC, per il tramite degli Istituti centrali, sono incoraggiati ad avviare un processo di standardizzazione sia per la rappresentazione di dati ricorrenti, indipendenti dallo specifico dominio applicativo, come per esempio i dati sulle persone, sulle organizzazioni pubbliche e private, sui luoghi e gli indirizzi usando le ontologie di OntoPiA³³ sia per la rappresentazione di dati più settoriali relativi a specifiche banche dati o a specifici procedimenti o per i dati pubblicati nella sezione «Amministrazione Trasparente».

È fondamentale, quindi, nella scelta delle ontologie da utilizzare, nell'ordine:

1. avvalersi di ontologie della rete OntopiA

³³ In particolare il MiC collabora ad OntoPiA con le ontologie: 1) Cultural-ON per la modellazione dei dati relativi agli Istituti della Cultura e agli eventi culturali; 2) ArCo, rete di ontologie per la modellazione dei beni culturali afferenti ai beni archeologici, architettonico-paesaggistici, demoetnoantropologici, fotografici, naturalistici, numismatici, scientifico-tecnologici, storico-artistici, strumenti musicali.

2. usare ontologie rilasciate come standard dal W3C
3. usare ontologie pubblicate e aggiornate da grandi Istituzioni (es Library on Congress, ICOM, Europeana)
4. usare altre ontologie che siano pubblicate su siti istituzionali, ben documentate e, preferibilmente, disponibili in almeno due lingue.

L'utilizzo di una rete di ontologie standard può facilitare la creazione di collegamenti tra dati, portando alla costruzione di una grande base di conoscenza dell'informazione del settore culturale da utilizzare per lo sviluppo di servizi nuovi e proattivi. Per una visione complessiva dell'architettura dell'informazione pensata per il settore pubblico si rimanda allo schema descritto da AgID all'indirizzo:

<https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/arch.html#id1>.

Quali sono gli standard di riferimento per il settore pubblico?

I principali standard di riferimento per l'architettura dell'informazione del settore pubblico derivano dalle esperienze maturate dagli esperti nel settore del Web Semantico, con la visione di trasformare il Web in un unico spazio informativo globale. Essi sono:

- RDF (Resource Description Framework)
- RDFS (RDF Schema)
- OWL (Web Ontology Language)
- SPARQL (SPARQL Protocol and RDF Query Language)
- SDMX (Statistical Data and Metadata eXchange)

Tali standard sono ampiamente documentati sul web e descritti anche nelle Linee guida AgID³⁴.

Cos'è il portale dati.gov.it?

Ai sensi dell'articolo 1 comma 8 del D.Lgs. 18 maggio 2015, n.102, il portale nazionale dei dati aperti (dati.gov.it) è l'unico riferimento per la

³⁴ <https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/arch.html#standard-di-riferimento>

documentazione e la ricerca di tutti i dati aperti della pubblica amministrazione. Esso, inoltre, è l'unico ad abilitare il colloquio con l'analogo portale europeo. Il portale nazionale dei dati aperti include i metadati, conformi al profilo DCAT-AP_IT, che descrivono i dati aperti delle amministrazioni. Le amministrazioni sono tenute pertanto a inserire e a mantenere aggiornati, attraverso le modalità di alimentazione previste dal catalogo, tali metadati. I dati primari, il cui riferimento è pubblicato sul portale nazionale, rimangono presso il titolare del dato che conserva la responsabilità della loro divulgazione a livello nazionale. I dati geografici devono essere documentati esclusivamente presso il Repertorio Nazionale dei Dati Territoriali (RNDDT) che, in maniera automatizzata, si occupa dell'allineamento con il portale nazionale dei dati.

Il MiC assicura il conferimento dei propri dati aperti a dati.gov.it attraverso il proprio Catalogo dei dati pubblicato sul sito <https://dati.beniculturali.it>.

Il portale dati.gov.it predispone i metadati per l'*harvesting* da parte del portale europeo e prevede una funzionalità di *harvesting* periodica (con frequenza settimanale) e automatizzata verso i cataloghi dei dati aperti delle altre amministrazioni.

Al fine di evitare duplicazioni di dati e di ridurre la complessità della raccolta centrale di informazioni sui dataset presenti nei cataloghi delle amministrazioni, si adotta un modello di *governance* del processo di alimentazione del catalogo nazionale dati.gov.it. Il modello di *governance* prevede di avvalersi dei principi di sussidiarietà verticale, già in precedenza menzionati. I meccanismi di alimentazione del portale nazionale abilitano, di fatto, una federazione tra portali di pubbliche amministrazioni⁷.

Cos'è il portale dati.beniculturali.it?

Il sito <https://dati.beniculturali.it> è il portale dove sono pubblicati i dati aperti del Ministero della cultura, la cui manutenzione è stata affidata dalla DG Organizzazione all'Istituto Centrale per il Catalogo e la Documentazione. Dal portale è possibile interrogare l'endpoint SPARQL dove sono accessibili i *linked open data* (LOD) prodotti dal MiC. I primi LOD sono stati pubblicati a partire dal 2014 rappresentano il frutto di un processo di cooperazione tra gli Istituti centrali e le Direzioni generali del MiC e collegano tra loro dataset provenienti da fonti diverse: banca dati dei Luoghi della cultura; anagrafiche di Archivi e Biblioteche; banca dati del Catalogo dei beni culturali; altre banche dati documentali e fotografiche. Allo stato attuale la piattaforma è essenzialmente un'interfaccia sperimentale machine-to-machine (m2m) che offre *linked open data* interrogabili via endpoint SPARQL.

