

Disponibilità di banche dati pubbliche in formato aperto

Nota metodologica

1. Introduzione

L'indicatore "Disponibilità di banche dati pubbliche in formato aperto" è inserito nell'Accordo di Partenariato 2014-2020 (AP) come *proxy* associata al Risultato Atteso 11.1 - *Aumento della trasparenza e interoperabilità, e dell'accesso ai dati pubblici*, ed è definito nell'AP come "Numero di banche dati pubbliche disponibili in formato aperto in percentuale sulle banche dati pubbliche di un paniere selezionato", con un valore obiettivo previsto nel 2023 (termine finale di misurazione dell'indicatore) pari al 90%.

La valorizzazione di tale indicatore, realizzata da AgID, è il frutto di una riflessione metodologica definita nell'ambito di un confronto tecnico, anche con l'Istat, coordinato dal NUV AP-DPCoe-PCM.

Il processo metodologico che ha condotto alla quantificazione del dato si è articolato in tre fasi:

1. Definizione del campo di osservazione, ovvero dei dataset da monitorare: le attività hanno visto la partecipazione di diversi attori sia della Pubblica Amministrazione che della società civile.
2. Definizione dell'indicatore che osserva il numero di dataset disponibili come dati di tipo aperto.
3. Calcolo dell'indicatore.

Si evidenzia che l'azione di aggiornamento del Paniere, che fornisce la base di calcolo del suddetto indicatore, rientra nella strategia complessiva del Piano triennale 2019-2021 (<https://www.agid.gov.it/it/agenzia/piano-triennale>) e rappresenta, anche con i successivi aggiornamenti del Piano, un obiettivo specifico in materia di dati di tipo aperto (open data), fermo restando quanto previsto dalla direttiva 1024/2019, recepita in Italia con il D. Lgs. n. 200/2021 che ha modificato il D. Lgs. n. 36/2006, e dagli atti di esecuzione della Commissione Europea riguardo ai "Dataset di elevato valore".

Dal 2019, per la sola componente nazionale del paniere, è stato variato l'insieme complessivo di dataset con l'introduzione di nuovi e l'esclusione di altri che sembravano ridondanti.

Complessivamente si registra un incremento sia a livello di componente nazionale che di componente regionale e la presenza di un numero significativo di dataset che appaiono non più aggiornati.

Di seguito si illustra la metodologia di calcolo dell'indicatore disaggregato a livello regionale.

2. Definizione dell'indicatore

La prima fase del processo ha riguardato la definizione del campo di osservazione, ovvero del set di banche dati pubbliche da includere nel paniere per il calcolo dell'indicatore. Al riguardo, la selezione è stata realizzata considerando sia le banche dati "nazionali" che quelle rese disponibili da Amministrazioni regionali e locali. Ciò ha condotto all'individuazione di due panieri distinti: uno *nazionale* e uno *regionale*.

Infatti, si è ritenuto metodologicamente corretto osservare sia le banche dati di responsabilità delle

Amministrazioni regionali e locali, che quelle gestite da Amministrazioni centrali, in considerazione della loro rilevanza nel panorama del patrimonio informativo pubblico.

Un prerequisito essenziale dei dataset osservati è la *proprietà dei dati*, ovvero la coincidenza tra il soggetto pubblico che rende disponibili e aperti i dati e l'Amministrazione titolare e responsabile della raccolta degli stessi. Pertanto, non sono considerati i dataset pubblicati su siti istituzionali di amministrazioni che provengono da fonti pubbliche diverse dagli enti stessi.

L'indicatore è dunque calcolato a livello regionale come:

$$I = \frac{\sum \text{dataset regionali aperti} + \text{dataset nazionali aperti}}{\sum \text{paniere totale dataset regionali} + \text{paniere totale dataset nazionali}}$$

3. Il set di banche dati del paniere nazionale

Il paniere nazionale è composto da una lista di 65 dataset riferiti a diversi ambiti tematici, così come definiti dalla classificazione dei temi dei dati del profilo di metadattazione italiano DCAT- AP_IT, scelti sulla base dei seguenti criteri:

1. dataset previsti dai piani di rilascio delle singole Amministrazioni centrali, come da elenco contenuto nelle Agende per la valorizzazione del patrimonio informativo pubblico degli anni 2013, 2014 e 2015;
2. dataset richiesti dalla società civile tramite consultazione pubblica ufficiale avviata dall'Agenzia per l'Italia Digitale alla fine del 2014;
3. dataset individuati nel piano d'azione italiano, in attuazione del “G8 Open Data Charter”;
4. dataset individuati nel terzo Piano d'azione dell'Open Government Partnership (OGP) nell'ambito dei temi Open Data e Trasparenza², definito anche con il coinvolgimento diretto di alcune associazioni della società civile e altri portatori di interesse;
5. alcune banche dati di interesse nazionale così come definite all'art. 60 del Codice dell'Amministrazione Digitale (ad esempio, la banca dati dei contratti pubblici).
6. dataset comunicati da alcune amministrazioni e ritenuti di rilevanza nazionale.

Dalla lista delle banche dati “nazionali”, sono stati esclusi:

- i dataset che contengono dati personali in quanto protetti dalle disposizioni legislative vigenti in materia di protezione dei dati personali (è il caso, ad esempio, della banca dati di interesse nazionale “Casellario giudiziale”);
- i dataset pubblicati in via occasionale per eventi e convegni;
- alcuni dataset richiesti ufficialmente mediante consultazione pubblica ma riferibili a enti territoriali quali comuni e regioni, per i quali è stato valutato di includere specifici dataset nel relativo paniere regionale (cfr. sezione successiva).

Ad ogni dataset presente nel paniere nazionale è stata associata una variabile dicotomica in ragione della valutazione del relativo stato (la variabile è pari a 1 se il dataset è aperto, 0 in caso contrario).

¹<http://sitiarcheologici.palazzochigi.it/www.funzionepubblica.it/febbraio%202016/www.funzionepubblica.gov.it/media/1109205/piano%20dazione%20g8%20open%20data%20%20it.pdf>

²<http://open.gov.it/2016/09/20/pubblicazione-terzo-nap/>

Per quanto riguarda la regionalizzazione del paniere nazionale, tutti i dataset considerati sono conteggiati per ciascuna regione.

4. Il set di banche dati del paniere regionale

Il paniere regionale è composto da una lista di 48 dataset osservati in ciascuna regione e riguardanti diversi ambiti tematici, secondo la classificazione dei temi dei dati del profilo di metadatazione italiano DCAT-AP_IT. In particolare, i dataset “regionali” sono stati inclusi nel paniere sulla base del seguente percorso:

1. Definizione di una prima lista di dataset (“paniere_v.0”) già disponibili in alcuni cataloghi aperti delle regioni (ad esempio Lombardia, Lazio, Umbria). In questa fase sono stati selezionati alcuni dataset di rilievo sul piano di un eventuale riutilizzo e quelli presenti in più di un catalogo, anche con l’obiettivo di incentivare, in prospettiva, un’apertura armonizzata degli stessi da parte di tutte le regioni attraverso l’adozione di modelli comuni;
2. Definizione di una seconda lista di dataset (“paniere_v.1”) realizzata da parte del gruppo di lavoro “Data & Open Data Management” istituito presso il Comitato di pilotaggio del Dipartimento della funzione pubblica relativamente agli interventi OT11 (Rafforzamento della capacità istituzionale e amministrativa) e OT2 (Attuazione dell’Agenda Digitale) dell’Accordo di Partenariato 2014-2020. I rappresentanti regionali, a partire dal “paniere_v.0”, hanno espresso un’indicazione circa le scelte di rilascio di dati aperti per il periodo 2014-2020;
3. Confronto e unione dei contenuti delle liste di cui ai punti 1 e 2 per il consolidamento di un unico paniere regionale (“paniere_v.2”), dal quale sono state ovviamente escluse non solo le eventuali banche dati presenti in entrambe le liste, ma anche quelle correttamente incluse nel paniere nazionale (ad esempio, quelle riguardanti i musei, le zone di produzione di vini DOP e IGP, gli eventi culturali, la carta d’uso dei suoli o i rifiuti).
4. Anche in questo caso, per ciascuna regione ad ogni dataset presente in “paniere_v.2” è stato associata una variabile dicotomica in ragione della valutazione del relativo stato (la variabile è pari a 1 se il dataset è aperto, 0 in caso contrario).

5. Il calcolo dell’indicatore

L’indicatore è calcolato come rapporto percentuale tra il numero di dataset pubblicati come dati di tipo aperto e il numero totale di dataset inclusi nei due panieri descritti nei paragrafi precedenti (nazionale e regionale).

Per “dataset di tipo aperto” si intende i dati pubblici che, in coerenza con quanto previsto dal CAD, sono resi disponibili sul web, anche in “bulk”, in formato aperto e con una licenza aperta (nei casi di licenza non specificata si è verificata l’applicabilità del principio *open data by default* così come definito all’articolo 52 comma 2 del Codice dell’Amministrazione Digitale).

Nel processo di monitoraggio finalizzato all’aggiornamento all’aggiornamento del Paniere alla data del 31/12/2021 non sono stati considerati i dataset che, seppur resi disponibili attraverso cataloghi open data, non risultano coerenti con i requisiti previsti dall’art. 1 del CAD (art. 1, co.1 lett l-bis e lett.l-ter) per quanto concerne l’accesso ai dati, la licenza e il formato dei dati stessi.

Inoltre, pur essendo riportati nel Paniere per darne evidenza, non sono stati considerati ai fini del calcolo dell’indicatore:

1. i dataset con copertura regionale ma riferibili solo a un numero limitato di comuni e/o province

della regione.

2. i dataset la cui copertura rispetto a quanto richiesto dalla descrizione dello stesso è limitata (e.g., tutti i pagamenti della regione vs un solo tipo di pagamenti).

Ciò stante, si intende proseguire nel processo di monitoraggio attraverso il diretto coinvolgimento delle amministrazioni titolari che potranno segnalare anomalie e richiedere eventuali modifiche/integrazioni.

6. Sviluppi

In prospettiva, ferma restando la modalità di calcolo dell'indicatore, sarà necessario valutare eventuali sviluppi in tema di monitoraggio tenendo conto:

1. delle eventuali indicazioni contenute nella strategia nazionale dati di cui all'articolo 50 ter del CAD come modificato dall'articolo 34 del D.L. n. 76/2020 convertito con modificazioni dalla Legge n. 120 del 2020;
2. delle indicazioni contenute nella Direttiva (UE) 2019/1024 del 20 giugno 2019, recepita in Italia con il D. Lgs. n. 200/2021 che ha modificato il D. Lgs. n. 36/2006, relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico, con particolare riferimento all'introduzione del concetto di "Dataset di elevato valore" (High Value Dataset);
3. collegato al punto precedente, delle disposizioni che saranno previste con gli atti di esecuzione della Commissione Europea riguardo ai citati "Dataset di elevato valore".