



Agenzia per l'Italia Digitale

Presidenza del Consiglio dei Ministri

Rapporto Annuale (anno 2016)

Disponibilità di banche dati pubbliche in formato aperto

Versione 0.5



Sommario

1.Introduzione.....	4
2.Definizione delle liste di dataset.....	6
2.1. Definizione della componente nazionale del paniere.....	6
2.2. Definizione della componente regionale del paniere.....	7
3.Individuazione di indicatori di misurazione.....	8
4.Monitoraggio e risultati.....	9
5.Accordo di Partenariato 2014-2020.....	11
6.Cronoprogramma e prossime azioni.....	13
Allegato I – modello per i dati di tipo aperto.....	14

Lista delle figure

Figura 1: Processo di monitoraggio adottato.....	4
Figura 2: Cronoprogramma monitoraggio per l'anno 2017.....	13

1. Introduzione

Il presente documento descrive l'attività di valorizzazione dell'indicatore “Disponibilità di banche dati pubbliche in formato aperto”, 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*, e definito nell'AP come “Numero di banche dati pubbliche disponibili in formato aperto in percentuale sulle banche dati pubbliche di un paniere selezionato”. La valorizzazione dell'indicatore, realizzata da AgID, è il frutto di una riflessione metodologica definita nell'ambito di un confronto tecnico, anche con l'Istat, coordinato dal NUVAP-DPCoe-PCM.

In particolare, il lavoro è stato avviato nel 2015 e ha portato alla definizione di un paniere di dataset costruito principalmente a partire dalle agende nazionali per la valorizzazione del patrimonio informativo pubblico 2013, 2014 e 2015, che includevano anche alcune richieste provenienti dalla società civile, e le banche dati di interesse nazionale ai sensi dell'articolo 60 del Codice dell'Amministrazione Digitale (CAD) nella sua formulazione disponibile a quel tempo.. Il paniere risultante e il relativo monitoraggio, che fotografava la situazione al 30 Aprile 2015, sono disponibili pubblicamente attraverso l'area di lavoro condivisa (ALC) “Data & Open Data Management”, attivata dal Comitato di pilotaggio per il coordinamento degli interventi OT11 OT2 realizzati nel quadro dell'AP Italia in tema di rafforzamento della capacità istituzionale e amministrativa (Obiettivo Tematico 11) e di attuazione dell'Agenda Digitale (Obiettivo Tematico 2)¹.

Nel corso del 2016 si è deciso di operare seguendo il processo di monitoraggio illustrato in Figura 1. Questo processo ha portato alla definizione di un nuovo paniere e di nuovi risultati di monitoraggio, descritti nel seguito del presente documento.

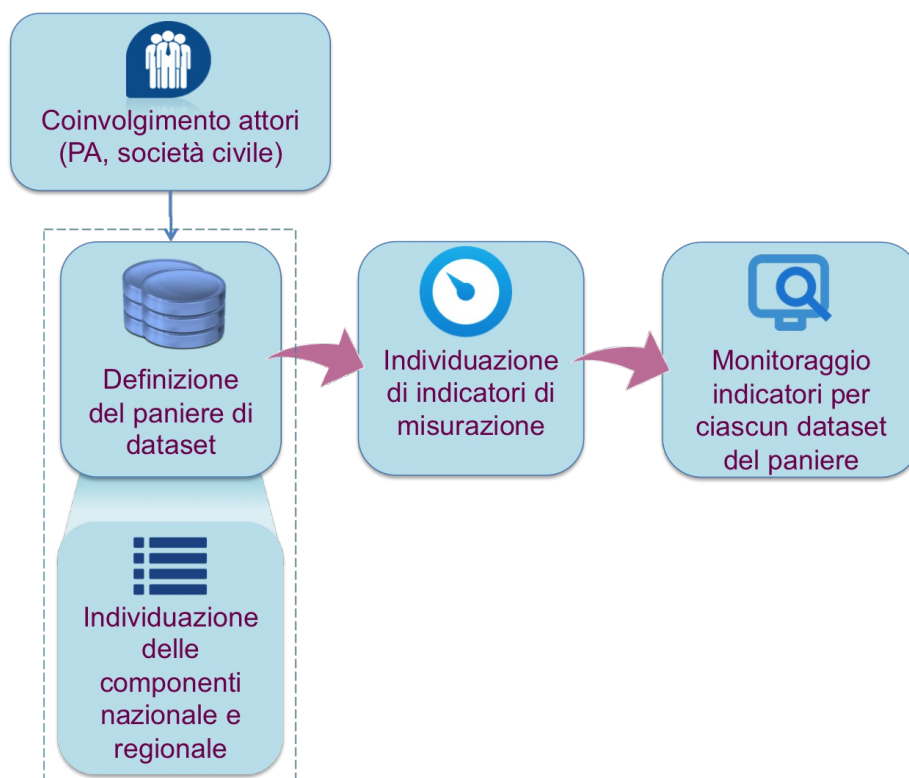


Figura 1: Processo di monitoraggio adottato

Il processo di monitoraggio complessivo illustrato in Figura 1 si sviluppa in tre fasi principali:

¹ http://network.ot11ot2.it/sites/default/files/paniereindicatori_20151231_agid.xlsx

- *definizione dei dataset da monitorare (campo di osservazione)*: questa fase ha visto il coinvolgimento anche di alcuni attori, sia nell'ambito della Pubblica Amministrazione, sia nell'ambito di iniziative di partecipazione della società civile come l'Open Government Partnership (OGP);
- *definizione degli indicatori di misurazione*: in questa fase è stato individuato un primo indicatore di misurazione dell'apertura di dati che rispecchia quanto previsto nell'ambito dell'AP; un secondo indicatore è stato solo preliminarmente discusso per misurare alcune caratteristiche fondamentali di qualità dei dati aperti e di uso/riuso dei dati;
- *monitoraggio* e quindi calcolo del suddetto indicatore.

Si evidenzia che il risultato della prima fase del processo è parte integrante della strategia complessiva italiana in materia di banche dati e dati aperti.

La presente nota metodologica descrive le varie fasi del processo di monitoraggio come illustrato in Figura 1 e le motivazioni che hanno determinato le scelte adottate nelle varie fasi. In particolare la sezione 2. descrive il metodo adottato per la selezione delle banche dati da monitorare; la sezione 3. discute dell'indicatore di misurazione utilizzato per effettuare il monitoraggio 2016 e la sezione 4. fornisce i risultati finali del monitoraggio sui panieri individuati. Infine, la sezione 6. conclude il presente documento illustrando un cronoprogramma con le azioni future da compiere, anche di concerto con le amministrazioni coinvolte.

2. Definizione delle liste di dataset

La prima fase del processo ha riguardato la scelta delle banche dati pubbliche da monitorare. La selezione è avvenuta seguendo diversi criteri rispetto al 2015 e ha portato alla formazione di un paniere che incorpora due componenti: una *componente nazionale* e una *componente regionale* in linea con le esigenze legate al monitoraggio dell'indicatore previsto nell'ambito dell'Accordo di Partenariato.

Accanto alla componente regionale di responsabilità delle Amministrazioni regionali e locali, sono state considerate anche banche dati gestite da Amministrazioni centrali (come peraltro avviato attraverso la definizione del paniere dell'anno 2015), tenuto conto sia della loro rilevanza nel panorama del patrimonio informativo pubblico, sia delle disposizioni contenute nel Codice dell'Amministrazione Digitale (art. 52) relativamente alla predisposizione del rapporto annuale per la valorizzazione del patrimonio informativo pubblico.

2.1. Definizione della componente nazionale del paniere

La componente nazionale del paniere è composta da una lista di 62 dataset suddivisi per varie tematiche, così come definite dalla classificazione dei temi dei dati del profilo di metadato italiano DCAT-AP_IT, e scelti sulla base dei seguenti criteri:

1. dataset “chiave” individuati nel piano d'azione italiano, in attuazione del “G8 Open Data Charter”²;
2. dataset “chiave” previsti dai piani di rilascio delle singole Amministrazioni centrali, come da elenco contenuto nelle Agende per la valorizzazione del patrimonio informatico pubblico degli anni 2013³, 2014⁴ e 2015 (quest'ultima non pubblicata);
3. dataset “chiave” richiesti dalla società civile tramite consultazione pubblica ufficiale avviata dall'Agenzia per l'Italia Digitale alla fine del 2014⁵;
4. dataset individuati nel terzo Piano d'azione dell'Open Government Partnership (OGP) nell'ambito dei temi Open Data e Trasparenza⁶, 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'articolo 60 del Codice dell'Amministrazione Digitale (e.g., banca dati dei contratti pubblici)

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 diverse da quella titolare.

Dalla lista delle banche dati, selezionate secondo i criteri sopra indicati, sono stati esclusi i dataset che contengono dati personali, (e.g., la banca dati di interesse nazionale “Casellario giudiziale”), i dataset richiesti dalla società civile in via generica in occasione di eventi e convegni, i dataset per i quali non poteva individuarsi in modo certo l'amministrazione titolare, e alcuni dataset richiesti ufficialmente mediante consultazione pubblica ma che essendo riferibili a enti territoriali sono stati inclusi nella componente regionale del paniere (sezione 2.2).

2 <http://sitiarcheologici.palazzochigi.it/www.funzionepubblica.it/febbraio>

[%202016/www.funzionepubblica.gov.it/media/1109205/piano%20dazione%20g8%20open%20data%20it.pdf](http://www.funzionepubblica.gov.it/media/1109205/piano%20dazione%20g8%20open%20data%20it.pdf)

3 <http://documentslide.com/documents/agenda-nazionale-per-la-valorizzazione-del-patrimonio-informativo-pubblico-agid-2013-bozza.html>

4 http://www.agid.gov.it/sites/default/files/documenti_indirizzo/agendanazionalepatrimoniopubblico2014.pdf

5 <http://www.agid.gov.it/notizie/2014/11/27/consultazioni-valorizzazione-patrimonio-informativo-pubblico>

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

La componente nazionale del paniere è pubblicata sul portale dei dati dati.gov.it.

2.2. Definizione della componente regionale del paniere

La componente prettamente regionale del paniere è composta da una lista di 48 dataset suddivisi per varie tematiche, così come definite dalla classificazione dei temi dei dati del profilo di metadattazione italiano DCAT-AP_IT, e scelti sulla base del seguente processo:

1. è stata definita una prima lista di dataset, con le relative descrizioni, partendo da quelli già disponibili in alcuni cataloghi aperti delle regioni (e.g., Lombardia, Lazio, Umbria). A tal riguardo, sono stati scelti dataset ritenuti di impatto per l'eventuale riutilizzo e dataset più ricorrenti nei vari cataloghi (in quest'ultimo caso anche al fine di incentivare, per il futuro, un'apertura armonizzata degli stessi da parte di tutte le regioni, attraverso la specifica di modelli comuni);
2. durante i lavori del gruppo “Data & Open Data Management”, attivato dal Comitato di pilotaggio per il coordinamento degli interventi OT11 OT2 realizzati nel quadro dell'AP Italia in tema di rafforzamento della capacità istituzionale e amministrativa (Obiettivo Tematico 11) e di attuazione dell'Agenda Digitale (Obiettivo Tematico 2), è stato richiesto alle regioni quali dataset intendessero rilasciare secondo il paradigma dei dati aperti nel periodo 2016-2020. Sulla base dei riscontri forniti da alcune regioni è stata definita un'ulteriore lista di dataset;
3. attraverso poi alcuni confronti sulle liste di dataset individuate nei punti 1 e 2 si è arrivati alla definizione di una prima lista complessiva di dataset regionali;
4. detta lista complessiva è stata verificata anche in funzione di alcune richieste provenienti da consultazioni pubbliche ufficiali, eliminando le duplicazioni riscontrate (e.g., alcune tipologie di dataset territoriali);
5. infine, dalla lista di dataset regionale suddetta sono stati eliminati alcuni dataset (e.g., musei, zone di produzione di vini DOP e IGP, eventi culturali, carta d'uso dei suoli, rifiuti, ecc.) in quanto già contemplati in dataset aperti da enti centrali e inseriti nella componente nazionale del paniere, come precedentemente descritto.

La componente regionale finale del paniere è pubblicata sul portale dei dati dati.gov.it.

3. Individuazione di indicatori di misurazione

Questa fase del processo di Figura 1 ha considerato diversi elementi che hanno portato alla definizione e discussione di due indicatori: *indicatore di apertura*, *indicatore di grado di apertura e qualità*.

Tenuto conto della descrizione dell'indicatore come fornita dall'Accordo di Partenariato, si è deciso, per l'anno 2016, di monitorare i dataset del paniere sia per la componente nazionale che per quella regionale. Ad ogni dataset delle due componenti 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). L'indicatore di apertura (a) è così calcolato:

$$(a) \text{ indicatore di apertura} = \#dataset \text{ aperti} / \#dataset \text{ monitorati} (\%)$$

dove per “dataset aperti” si è inteso dati pubblici resi disponibili e accessibili sul web, anche in “bulk”, 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) e secondo uno qualsiasi dei livelli previsti dal modello per i dati aperti delle linee guida AgID per la valorizzazione del patrimonio informativo pubblico, riportato per facilità di lettura nell'Allegato I del presente documento. Alla luce di tale considerazione, anche dataset di livello 1 e 2 del modello dei dati aperti concorrono al calcolo dell'indicatore di apertura. Questa scelta è motivata dal fatto che solo nelle linee guida per la valorizzazione del patrimonio informativo pubblico (anno 2016) sono state incluse raccomandazioni più stringenti in merito alla pubblicazione di *sol*i dati strutturati con licenza aperta ma in formato proprietario. Pertanto, per consentire a tutte le amministrazioni/regioni di adeguarsi alle nuove indicazioni, si è ritenuto di procedere secondo i principi suddetti.

Tuttavia, come riportato nel crono programma delle attività (sezione 6.), è stato ipotizzato che, a partire dal 2017, il monitoraggio tenga conto anche di un altro indicatore, chiamato indicatore di grado di apertura e qualità (indicatore (b)). Ove tale ipotesi venisse confermata, si terranno in considerazione non solo l'apertura dei dataset ma anche elementi di misurazione della qualità dei dati e dei metadati, e di uso/riuso dei dataset stessi. Sempre a partire dal 2017, anche al fine di avere un riscontro diretto da parte delle amministrazioni competenti, il monitoraggio potrà essere effettuato in base ai risultati di un questionario da sottoporre alle amministrazioni, preventivamente informate, attraverso il portale dati.gov.it, e procedendo poi a eventuali verifiche a campione.

4. Monitoraggio e risultati

Nel 2015, risultavano aperti circa il 40% dei dataset del paniere allora identificato.

Per quanto riguarda invece il 2016, i risultati ottenuti per l'indicatore (a) relativamente alla componente nazionale del paniere monitorata fino a Novembre 2016 sono i seguenti:

	Apertura anno 2016
Indicatore (a)	59,68% (37dataset aperti su 62)

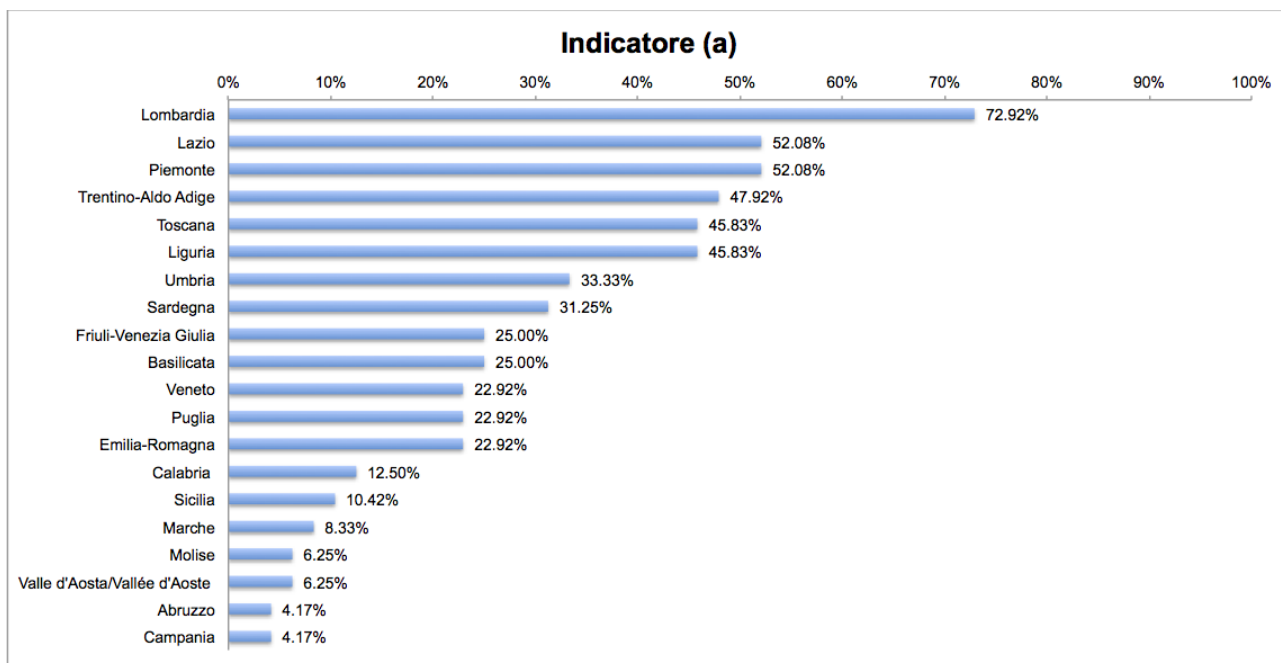
Nel caso del paniere regionale, si noti che nell'effettuare il monitoraggio le seguenti fonti sono state considerate per le varie regioni:

- cataloghi di dati aperti o pagine del sito web istituzionale della pubblica amministrazione/regione relative a dati aperti;
- siti web delle arpa regionali per alcune tipologie di dataset (e.g., dati sui sensori meteo, dati meteo), dove la descrizione dei dataset stessi indicava di riferirsi a tali fonti;
- nel caso della regione Trentino Alto Adige, il catalogo dei dati aperti della provincia autonoma di Trento e quello della provincia autonoma di Bolzano;
- nei casi delle regioni Campania e Sicilia che non hanno, alla data del presente monitoraggio (30 novembre 2016), un catalogo di dati aperti o pagine del sito istituzionale con dati aperti facilmente raggiungibili i “non portali dei dati aperti” creati dalla spinta di comunità locali attive in questo settore. I “non-portali” rappresentano un unico punto di accesso a un insieme di dati aperti e non aperti delle due regioni e sono stati un utilissimo strumento di orientamento nell'identificazione di dataset effettivamente aperti o con una licenza non specificata per la quale applicare il principio dell'open data by default, come definito nell'art. 52 del Codice dell'Amministrazione Digitale.

I risultati per regione, in ordine decrescente, per i 48 dataset della componente regionale del paniere sono rappresentati nella seguente tabella.

Indicatore (a)	Regione
72,92%	Lombardia
52,08%	Lazio
52,08%	Piemonte
47,92%	Trentino-Aldo Adige (Provincia Autonoma di Trento, Provincia Autonoma di Bolzano)
45,83%	Toscana
45,83%	Liguria
33,33%	Umbria
31,25%	Sardegna
25,00%	Friuli-Venezia Giulia
25,00%	Basilicata
22,92%	Veneto

22,92%	Puglia
22,92%	Emilia-Romagna
12,50%	Calabria
10,42%	Sicilia
8,33%	Marche
6,25%	Molise
6,25%	Valle d'Aosta/Vallée d'Aoste
4,17%	Abruzzo
4,17%	Campania



5. Accordo di Partenariato 2014-2020

I risultati precedenti sono stati poi utilizzati per il calcolo complessivo dell'indicatore dell'Accordo di Partenariato, sulla base di una riflessione metodologica definita nell'ambito di un confronto tecnico, anche con l'Istat, coordinato dal NUVAP-DPCoe-PCM.

A tal fine, l'indicatore è calcolato a livello regionale come:

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

dove *I* rappresenta il rapporto percentuale tra il numero di dataset pubblicati in formato aperto e il numero totale di dataset inclusi nelle due componenti, nazionale e regionale, del paniere. Per tale calcolo si è provveduto a regionalizzazione la componente nazionale del paniere: tutti i dataset considerati sono conteggiati per ciascuna regione. I risultati ottenuti sono i seguenti:

Indicatore <i>I</i>	Regione
65,5%	Lombardia
56,4%	Lazio
56,4%	Piemonte
54,5%	Trentino-Aldo Adige (Provincia Autonoma di Trento, Provincia Autonoma di Bolzano)
53,6%	Liguria
53,6%	Toscana
48,2%	Umbria
47,3%	Sardegna
44,5%	Basilicata
44,5%	Friuli Venezia Giulia
43,6%	Emilia-Romagna
43,6%	Puglia
43,6%	Veneto
39,1%	Calabria
38,2%	Sicilia
37,3%	Marche
36,4%	Molise
36,4%	Valle d'Aosta/Vallée d'Aoste
35,5%	Abruzzo
35,5%	Campania
	Dato complessivo

45,7%	Italia
	Ciclo di programmazione F. S. 2014-20
49,5%	Regioni più sviluppate
39,7%	Regioni in transizione
40,2%	Regioni meno sviluppate

6. Cronoprogramma e prossime azioni

Come precedentemente discusso, per l'anno 2017, si prevede che il monitoraggio venga effettuato mediante il coinvolgimento diretto delle amministrazioni competenti peraltro sulla base di un altro indicatore (indicatore (b)) in fase di definizione. Le azioni di monitoraggio sono schematizzate, in una prima ipotesi di cronoprogramma, in Figura 2.

Si sottolinea che il paniere dei dataset potrà essere rivisto sulla base delle esigenze operative delle amministrazioni e tenendo comunque conto dei diversi impegni assunti nel contesto delle iniziative dell'Open Government Partnership.

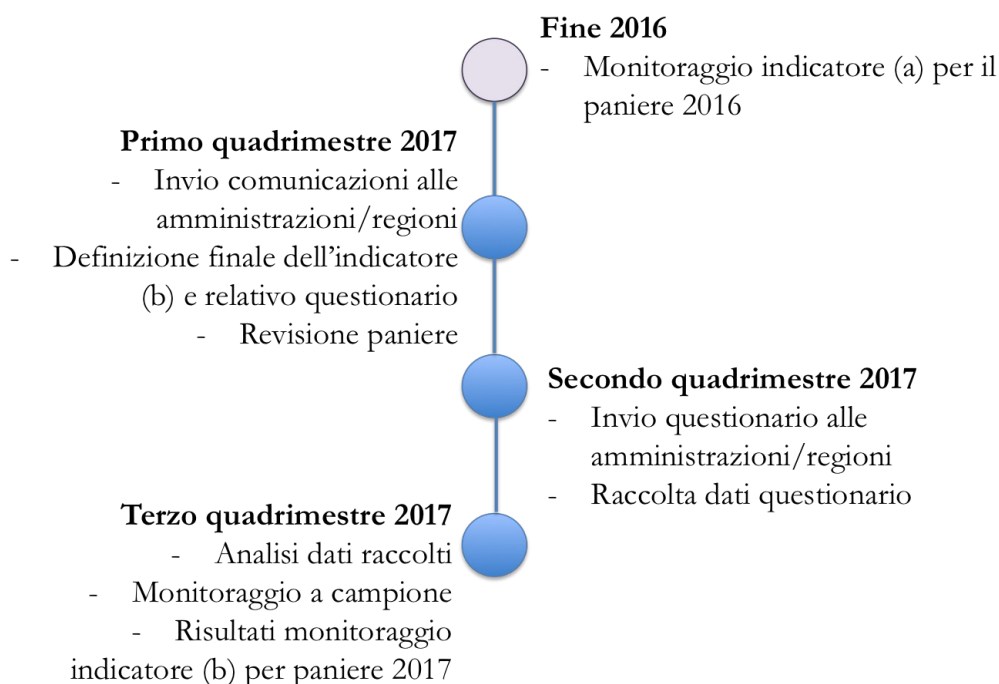


Figura 2: Cronoprogramma monitoraggio per l'anno 2017

Allegato I – modello per i dati di tipo aperto

