



## ALLEGATO A alla Dgr n. 2433 del 20 dicembre 2013

*Linee guida  
per il processo di standardizzazione  
delle banche dati*

### **Premessa**

Il presente documento, redatto nell'ambito del processo di semplificazione amministrativa previsto dalla DGR 1419 del 31.7.2012, contiene le linee guida di carattere generale per la creazione e la gestione di archivi amministrativi regionali in forma tendenzialmente standardizzata.

Si ritiene che la standardizzazione possa contribuire in modo significativo alla semplificazione amministrativa, consentendo di fornire risposte più efficaci, pertinenti e tempestive ai soggetti che le richiedono e, al tempo stesso, riducendo l'onere informativo a carico della collettività.

Alla base del documento vi è consapevolezza della molteplicità e dell'eterogeneità, formale e sostanziale, degli archivi amministrativi attualmente esistenti presso gli uffici regionali e del potenziale informativo e conoscitivo in essi racchiuso, considerati sia individualmente che in connessione con altre banche dati.

L'obiettivo delle linee guida è quello di sensibilizzare gli uffici regionali nonché gli Enti strumentali/Aziende afferenti alla Regione ad adottare livelli standard di qualità nella costruzione e nella gestione degli archivi amministrativi e di promuovere prassi virtuose per la valorizzazione, l'interscambiabilità e l'integrazione del patrimonio informativo regionale, in ossequio ai principi di economicità ed efficienza dell'azione amministrativa.

### **1. Principi generali**

L'amministrazione regionale dispone di una considerevole mole di dati in archivi formati e gestiti, con vario livello di informatizzazione, per finalità amministrative. I dati raccolti in questi archivi, tuttavia, non sempre sono adatti ad essere elaborati o riutilizzati per scopi diversi da quelli per cui sono stati raccolti. In tal modo, viene disperso un patrimonio informativo che andrebbe invece opportunamente valorizzato, al duplice scopo di abbattere gli oneri informativi a carico di cittadini ed imprese e di ridimensionare il "costo conoscitivo" sopportato dall'amministrazione.

Infatti, ogni archivio amministrativo, per quanto generato da esigenze di carattere gestionale, è una potenziale e preziosa fonte di informazioni; esso può diventare uno strumento conoscitivo nella misura in cui le informazioni registrate siano strutturate e collegate tra loro secondo un modello logico che ne consenta l'utilizzo a fini conoscitivi: in tale ottica, nel presente documento la locuzione "archivio amministrativo" è sostanzialmente equivalente a quella di "banca dati".

Per questo motivo, è fondamentale che l'archivio amministrativo sia concepito, sin dalla fase di creazione o ristrutturazione, come una cellula del più ampio organismo del sistema informativo regionale, e

quindi in funzione al perseguimento degli obiettivi istituzionali dell'intera amministrazione regionale e non solo dell'ufficio che lo gestisce.

Il valore informativo di un archivio amministrativo è direttamente proporzionale alla qualità della sua struttura (fattore intrinseco) e alla capacità di integrarsi e dialogare in modo efficace con altri archivi amministrativi (fattore estrinseco, o di sistema).

L'obiettivo dell'integrazione tra le banche dati richiede un livello minimo comune di standardizzazione in termini di strutturazione informatica e di gestione delle informazioni: l'integrazione tra le potenziali fonti informative sarà tanto più efficiente quanto più alto è il livello di standardizzazione dell'intero processo di acquisizione e registrazione dei dati, a partire dalla modulistica adottata.

La banca dati va, quindi, gestita e aggiornata secondo una metodologia affidabile, intelligente, efficiente e sostenibile, e tendenzialmente in una prospettiva temporale di lungo termine, poiché spesso il suo valore informativo trascende quello gestionale per cui è nata.

La corretta gestione degli archivi amministrativi implica l'osservanza della normativa in materia di dati delle pubbliche amministrazioni - e in primo luogo delle disposizioni sull'amministrazione digitale (d.lgs. 82/2005), sulla tutela della riservatezza dei dati personali (d.lgs. 196/2003), sugli *open data* (l. 221/2012) e sulla trasparenza (d.lgs. 33/2013) - e il necessario coordinamento con le iniziative regionali di semplificazione e *open government*, nel cui ambito la gestione delle informazioni costituisce un punto essenziale.

In una società sempre più fondata sulla conoscenza, il dato di qualità ha un valore anche sotto il profilo economico; pertanto, occorre prestare particolare attenzione agli aspetti giuridici connessi alla titolarità e alla disponibilità, anche futura, dei dati, allo scopo di curare al meglio gli interessi, anche patrimoniali, dell'Amministrazione regionale e della collettività.

## **2. Costruzione e gestione della banca dati**

La costruzione di una banca dati di qualità implica un presidio costante del processo di formazione dell'informazione nelle sue diverse fasi (raccolta/acquisizione, registrazione, strutturazione ed elaborazione dei dati), anche al fine di estenderne il più possibile l'utilizzabilità a fini informativi e l'integrabilità con altre fonti.

La fase di raccolta/acquisizione del dato è strategica e dev'essere pianificata adeguatamente: in particolare, occorre definire esplicitamente sin dall'inizio:

- le finalità per le quali sono necessari i dati; per fare ciò, è necessario muovere da un attento esame della normativa alla base del flusso informativo;
- le fonti dei dati pertinenti al processo eventualmente già disponibili;
- il campo di osservazione e le caratteristiche/variabili da rilevare;

Tali operazioni sono soggette al rispetto dei principi di necessità, pertinenza, non eccedenza e riduzione dell'onere informativo.

Conseguentemente:

- vanno raccolti tutti e solo i dati che effettivamente servono allo scopo gestionale e conoscitivo predefinito;

- risulta indispensabile, ai fini dell'integrazione tra banche dati, adottare strumenti di codifica per le variabili adottate (cfr. *infra*);
- qualora si usi della modulistica, è fondamentale la predisposizione di modelli redatti in modo chiaro e univoco, con linguaggio non burocratico e contenenti l'esatta definizione dei fenomeni da accertare, in modo da ridurre ambiguità ed errori; ove possibile, vanno adottate campiture standardizzate per la raccolta delle informazioni comuni (es. dati anagrafici).
- corrispondentemente, è necessario curare una corretta informazione e formazione di tutti i soggetti coinvolti nel processo di raccolta, con particolare attenzione nei confronti dei soggetti che, nella varie fasi, inseriscono le informazioni nella banca dati, anche per creare un clima di trasparenza e condivisione della rilevanza delle operazioni di registrazione dei dati;
- occorre preventivamente verificare l'esistenza, presso l'amministrazione regionale o altre pubbliche amministrazioni, di archivi recanti i medesimi dati richiesti, allo scopo di evitare duplicazioni antieconomiche, e privilegiare l'acquisizione d'ufficio dei dati registrati in banche dati certificate;
- in ogni caso, è opportuno abbattere, per quanto possibile, l'onere informativo a carico del soggetto rispondente (cittadino, impresa, P.A.), evitando di acquisire informazioni già in possesso dell'Amministrazione o comunque reperibili con altre modalità;
- va posta adeguata attenzione alle date da collegare ai dati o a gruppi di dati, in quanto la corretta datazione delle informazioni permetterà in seguito la collocazione temporale dei fenomeni e la visione diacronica degli stessi.
- nella scelta di modalità e strumenti per l'archiviazione dei dati, oltre a parametri quali la numerosità delle informazioni da gestire, l'episodicità o ripetitività nel tempo, la disponibilità di risorse, va tenuto conto anche dell'esigenza, di interesse generale, di potenziale riutilizzo dei dati a fini conoscitivi;
- nei casi in cui una o più fasi di strutturazione o gestione della banca dati vengano conferite o delegate, a qualunque titolo, a soggetti esterni all'Amministrazione regionale (ad es., contratti con società esterne, convenzioni con società regionali o altri enti pubblici, ecc...), occorre garantire con apposite previsioni contrattuali e convenzionali che gli interventi esterni rispettino i canoni informatici del sistema informativo regionale e che l'Amministrazione regionale abbia la disponibilità dei dati e possa accedervi liberamente.

Inoltre, in tutte le fasi di formazione dell'archivio, e specialmente durante la raccolta, occorre controllare periodicamente la correttezza dei dati acquisiti, per evitare errori comuni quali lacunosità o incoerenza dei dati raccolti, distorsioni volontarie delle dichiarazioni, discrasie temporali nelle registrazioni, patologie organizzative o procedurali.

### **3. Informazioni sugli archivi: le codifiche e i metadati**

Un aspetto fondamentale per un corretto sviluppo e funzionamento della banca dati è la predisposizione e l'aggiornamento delle informazioni sugli archivi, ovvero delle codifiche e dei metadati.

Tali informazioni costituiscono la "carta d'identità" degli archivi, perché descrivono le variabili e le relative modalità di raccolta e di gestione lungo tutta la vita dell'archivio stesso.

A tale proposito, occorre:

- se attinenti, adottare classificazioni e codifiche ufficiali già esistenti e in uso a livello internazionale, europeo e nazionale, privilegiando, ove possibile, il livello territoriale più ampio;

- se non sono disponibili codifiche ufficiali per descrivere il fenomeno di interesse ed è quindi necessario creare nuove classificazioni, rispettare criteri di base di chiarezza, trasparenza e univocità, evitando sovrapposizioni di definizioni e ambiguità; le nuove classificazioni e codifiche così create potranno essere approvate e condivise a livello regionale e divenire standard ufficiali regionali;

- inserire sempre, ove possibile, codici standard identificativi delle unità rilevate (es. codice fiscale della persona fisica o giuridica, partita IVA delle imprese, ecc...), al fine di costituire una connessione con altri archivi che prevedono i medesimi codici e consentire un'integrazione lineare.

I metadati descrivono i set di dati e rappresentano le informazioni e i contenuti delle caratteristiche tecniche, scientifiche e funzionali dei dataset stessi al fine di rendere semplice ed univoca l'identificazione. A titolo di esempio, rientrano tra i metadati le indicazioni sulla fonte di origine dei dati, il riferimento temporale, il livello territoriale, la definizione del calcolo degli indicatori presenti nell'archivio, le caratteristiche tecnico informatiche dei dati, le caratteristiche georeferenziate, se esistenti.

Tali strumenti consentono inoltre di ricostruire diacronicamente la vita dell'archivio, registrando nel corso del tempo le vicende che lo hanno caratterizzato ( ad es., eventuali modifiche nelle definizioni e nelle descrizioni di fenomeni: aggiunta o riduzione di variabili, ecc...). Tali informazioni infatti consentono di seguire la vita dell'archivio nello spazio e nel tempo, garantendo che i dati in esso contenuti, possano essere utilizzati e confrontati in modo ragionato e consapevole anche se risalenti nel tempo.

#### **4. Interoperabilità**

Con il termine interoperabilità si intende in generale la capacità di due o più soggetti di scambiarsi informazioni o servizi in modo più o meno automatico, con il minor numero possibile di errori, privilegiando affidabilità, sicurezza e ottimizzazione delle risorse.

Obiettivo dell'interoperabilità è dunque facilitare l'interazione fra sistemi differenti, nonché lo scambio e il riutilizzo delle informazioni anche fra sistemi informativi non omogenei (sia per software che per hardware).

Negli ultimi anni il concetto di interoperabilità è si andato sempre più diffondendo nelle pubbliche amministrazioni, dove la necessità di far dialogare sistemi informativi diversi è divenuta sempre più pressante, anche a causa di obblighi di Legge specificamente introdotti in tal senso.

Nel nostro caso, l'applicazione del concetto di interoperabilità alle banche dati e la sua utilità sono facilmente intuibili. Basti pensare a quante volte archivi ricchi di informazioni non sono collegabili tra loro a causa di una mancanza di chiavi identificative comuni, oppure per una scarsa qualità/aggiornamento delle informazioni che rende incerta qualsiasi conclusione assunta sulla base dell'analisi di tali informazioni.

Per interoperare è necessario che i diversi componenti del sistema, nel nostro caso le diverse banche dati, parlino tutti lo stesso linguaggio e ci siano delle regole a priori sulle modalità per eseguire ogni passaggio:

- all'interno di ciascun Ente, dove i diversi uffici devono dialogare fluentemente per via telematica, con procedure condivise ed accettate da tutti, senza duplicazioni, con una chiara attribuzione dei compiti, delle responsabilità e dei privilegi d'accesso alle informazioni;

- a livello regionale, dove i sistemi ed i servizi attivati dai vari Enti devono essere integrati, consentendo la piena cooperazione, per arrivare ad offrire agli utenti finali servizi completi e realmente integrati;

- a livello nazionale, rispetto al quale devono essere stabiliti degli standard comuni di riferimento, che consentano ad un Ente di dialogare con un qualsiasi altro Ente, ottimizzando le risorse esistenti, e a ciascun utente finale di ricevere veri servizi amministrativi on line, in grado di adattarsi velocemente al mutare delle esigenze.

Concettualmente l'interoperabilità può essere considerata su quattro livelli:

- tecnico: tale livello concerne le questioni tecniche di collegamento tra sistemi, la definizione delle interfacce e dei livelli di sicurezza, il formato dei dati e i protocolli, comprese le telecomunicazioni;

- semantico: si occupa di assicurare che il significato esatto delle informazioni scambiate sia comprensibile da qualsiasi altra applicazione, anche non pensata inizialmente per quel determinato scopo. Può assumere la forma di tassonomie di riferimento, schemi, elenchi di codici, dizionari di dati settoriali, biblioteche, ecc.

- gestionale o organizzativo: al fine di favorire il dialogo tra gli uffici coinvolti nel processo di interoperabilità deve essere garantito un livello adeguato e comune di servizio in termini di tempistiche, grado di aggiornamento, service desk, referenti..

- giuridico: le informazioni scambiate devono essere fruibili per le finalità delle organizzazioni. Il fatto che un'informazione sia disponibile non implica che questa sia usabile per fornire servizi. Solo un riconoscimento giuridico della possibilità di usare quel dato consente di utilizzarlo per l'erogazione dei servizi.

Nella costituzione di una banca dati è necessario che tutti gli aspetti sopra illustrati vengano attentamente esaminati. Il costo per la raccolta e la gestione delle informazioni infatti è considerevole ed accorgersi alla fine del processo che i dati raccolti sono inutilizzabili in tutto o in parte può condurre a spiacevoli conseguenze e a impatti molto negativi: costose bonifiche, ripetizione di indagini con ulteriore aggravio sui soggetti che forniscono i dati, ecc.

## **5. Qualità dell'archivio**

La standardizzazione e l'integrazione delle banche dati sono funzionalmente connesse per la realizzazione di un sistema informativo di qualità, poiché la standardizzazione delle banche dati mira a migliorare il livello qualitativo degli archivi e che l'integrazione tra archivi è utilmente realizzabile solo in presenza di archivi costruiti secondo criteri di qualità.

In tal senso, ai fini della diffusione e condivisione dell'archivio, i dati in esso contenuti devono soddisfare precisi requisiti di qualità quali:

- pertinenza: capacità dell'informazione di soddisfare le esigenze conoscitive degli operatori e degli utenti;

- accuratezza e attendibilità: livello di precisione nella misurazione del fenomeno oggetto di interesse;

- tempestività e puntualità: riduzione dei tempi intercorrenti tra l'evento descritto dai dati e la disponibilità degli stessi;

- confrontabilità: possibilità di confrontare nel tempo e nello spazio le informazioni del fenomeno di interesse. Il grado di confrontabilità è influenzato, oltre che dalle modificazioni concettuali che possono intervenire nel tempo e nello spazio, anche da cambiamenti intervenuti nelle definizioni e/o nelle caratteristiche operative adottate dal processo di produzione dell'informazione. Questo richiede un

mantenimento dell'archivio storico e un'adeguata comunicazione delle innovazioni e dei cambiamenti nelle modalità di alimentazione e tenuta degli archivi;

- accessibilità e chiarezza: trasparenza e semplicità per l'utente di reperire, acquisire e comprendere le informazioni disponibili in relazione alle proprie finalità. I dati devono essere disponibili e accessibili con imparzialità, con i relativi metadati e le necessarie istruzioni;
- integrabilità dei dati: esistenza e correttezza delle chiavi di identificazione condivise con altri archivi e delle chiavi di raccordo, utilizzo di definizioni, classificazioni e standard comuni;

## **6. Sicurezza informatica**

La sicurezza non è un prodotto, ma un processo e le banche dati non fanno eccezione a questa regola. La sicurezza degli archivi deve permeare tutte le fasi del ciclo di vita, dalla progettazione alla gestione quotidiana.

Le minacce a cui sono soggetti gli archivi informatici sono la perdita di integrità, di disponibilità e di confidenzialità. Per proteggere i database da queste minacce, possono essere implementate le seguenti misure di sicurezza:

- controllo degli accessi: il meccanismo di sicurezza di un sistema di gestione delle basi di dati deve prevedere forme di restrizione di accesso ai dati. Tali restrizioni si basano sulla creazione di account (utenti che accedono al dato), sull'attribuzione e revoca dei privilegi e sull'assegnazione dei livelli di sicurezza (autorizzazione);
- controllo di flusso: esso disciplina il modo in cui i dati raggiungono i soggetti autorizzati. Ad esempio il servizio, il numero di porta ed il trasporto (TCP o UDP) definiscono i canali leciti all'interno dei quali deve avvenire la comunicazione con i soggetti autorizzati. Eventuali comunicazioni al di fuori di tali canali devono essere inibite;
- crittografia: si tratta di un mezzo per proteggere i dati da accessi non autorizzati, o meglio per rendere intellegibili i dati ai soli soggetti autorizzati. La crittografia è l'approccio che di prassi è adottato per proteggere i dati acceduti da servizi accessibili da Internet e/o per proteggere i dati critici per un'organizzazione o sensibili sotto il profilo della privacy;
- backup: si tratta una misura di sicurezza volta principalmente a garantire la disponibilità dei dati e consiste nell'effettuare copie della base informativa da conservare anche presso un sito alternativo a quello che ospita l'archivio; le copie dei dati possono essere utilizzate per ripristinare in uno stato stabile e consistente il database i cui dati sono stati corrotti.

## **7. Organizzazione e strumenti**

Posto che la standardizzazione e l'impiego delle banche dati a fini conoscitivi sono processi, anche culturali, di lungo periodo, che necessitano di uno specifico supporto tecnico, appare opportuno dare continuità all'azione di standardizzazione delineata nel presente documento supportando l'Amministrazione con un complesso organico di specifici strumenti e iniziative, tra cui:

1. la creazione di un'area nelle rete Intranet regionale per promuovere il dialogo con gli uffici regionali interessati e la condivisione di informazioni, documenti operativi e *best practices* sulla creazione e gestione delle banche dati;

2. la costituzione di un Gruppo tecnico di coordinamento - composto dalle strutture competenti in materia di statistica, informatica e sistemi geografici, nonché dalle strutture di volta in volta interessate alla tematica specifica - per il supporto tecnico-scientifico degli uffici regionali nei processi di creazione, gestione, ristrutturazione e informatizzazione delle banche dati;

3. lo sviluppo di *softwares* specifici per contenere i metadati dei flussi informativi, documentare i cambiamenti della modulistica e degli archivi per consentire il confronto temporale, allo scopo di ottenere, per quanto possibile, una “banca dati delle banche dati regionali”;

4. l'organizzazione di eventi formativi e di sensibilizzazione.

Naturalmente, il successo di tale progetto richiede il contributo attivo di tutti gli uffici dell'Amministrazione regionale, i quali, pertanto, sono invitati a comunicare al Gruppo tecnico di coordinamento sopra citato ogni iniziativa che implichi una o più delle seguenti attività:

1. creazione, modifica o aggiornamento di una banca dati o della relativa modulistica;
2. creazione e gestione dell'archivio storico di una banca dati;
3. selezione delle fonti informative ed impiego di codifiche ufficiali e metadati;
4. strutturazione logico-informatica della banca dati.