

Redazione preliminare

I dati pubblici: linee guida per l'accesso, la comunicazione e la diffusione

8 febbraio 2002

Indice

1. Conoscibilità, sicurezza e usabilità dei dati
 2. Classificazione dei dati pubblici e loro rappresentazioni
 3. Modalità di attuazione della conoscibilità e soggetti interessati
 4. Qualità dei dati utili per la conoscibilità
 5. Requisiti di qualità
 - 5.1 Requisiti per la sicurezza
 - 5.2 Requisiti per l'usabilità
 - 5.3 Requisiti per le altre qualità
 6. Problematiche connesse alla distribuzione
 7. Raccomandazioni
- Appendice 1 : glossario

I dati pubblici: linee guida per la conoscibilità, l'accesso, la comunicazione e la diffusione

1. Conoscibilità, sicurezza e usabilità dei dati

I progetti di governo elettronico pianificati nei paesi dell'Unione europea hanno come linea direttrice quella di trasformare, attraverso l'uso delle tecnologie dell'informazione e della comunicazione (information and communication technology, ICT), le relazioni interne ed esterne del settore pubblico: relazioni tra soggetti pubblici e cittadini, tra soggetti pubblici e imprese o altri operatori privati e tra soggetti pubblici stessi, con il fine di migliorare l'erogazione dei servizi e la partecipazione della società civile alla vita democratica. L'impiego delle tecnologie è strumentale; il vero obiettivo è la trasformazione delle relazioni fra soggetti, che richiede un rilevante cambiamento culturale in particolare per i soggetti pubblici tenuti, sotto precisi requisiti e condizioni, a rendere conoscibili i dati in loro possesso, garantendone la qualità e l'aggiornamento.

I soggetti pubblici, in virtù dei loro compiti istituzionali, raccolgono e trattano grandi quantità di informazioni, codificate in forma di dati, sui cittadini, le imprese, le istituzioni, il territorio e i principali fenomeni della vita del paese. La *conoscibilità* di tali dati è costituita dall'insieme di regole che disciplinano la fruibilità in favore dei soggetti interessati da parte dei soggetti pubblici che li raccolgono e li trattano. La conoscibilità comporta due fondamentali *qualità* dei dati: la *sicurezza*, intesa come insieme delle misure tese ad assicurare l'accesso ai soli dati conoscibili a un soggetto interessato, e l'*usabilità*, intesa come la facilità con cui un soggetto interessato ed abilitato a conoscere il dato riesce ad accedervi con le tecnologie disponibili, tenuto conto della sua situazione fisica, psichica e culturale.

Un dato conoscibile può essere reso noto al soggetto interessato mediante tre *modalità di scambio*: *accesso*, *comunicazione*, *diffusione*. L'accesso permette al soggetto interessato di fruire direttamente del dato; la comunicazione consiste nel far pervenire il dato ad uno o più destinatari abilitati e predeterminati; la diffusione consiste nel rendere i dati disponibili ad una platea indeterminata di soggetti, anche tramite la loro pubblicazione, in forma tradizionale o su internet.

La crescente disponibilità di dati in formato digitale e la diffusione delle tecnologie internet impongono un riesame della problematica relativa alla conoscibilità, a causa dell'aumentata accessibilità, del venir meno di limiti territoriali e della diminuzione del costo. Le nuove tecnologie, peraltro, se da un lato favoriscono la usabilità dei dati, con l'uso di interfacce di comprensione sempre più intuitiva, dall'altro impongono obblighi di protezione per chi li detiene e sollevano problemi di discriminazione tecnologica verso gli utenti e di concorrenza tra la pubblica amministrazione e il settore privato.

Per le considerazioni appena esposte, l'Autorità per l'informatica nella pubblica amministrazione ha ritenuto di dover riesaminare il tema della conoscibilità dei dati posseduti da soggetti pubblici, nel contesto delle trasformazioni indotte dall'ICT, con il duplice intento di fornire indicazioni alle pubbliche amministrazioni sulla tematica e formulare raccomandazioni al legislatore per l'elaborazione di nuove norme in materia di conoscibilità dei dati pubblici.

Il presente documento rappresenta il risultato di tale attività. Viene inizialmente affrontata la problematica dei dati pubblici e sono fornite alcune classificazioni rilevanti; vengono individuate le modalità di attuazione della conoscibilità e le tipologie di soggetti interessati a fruire dei dati pubblici ovvero a distribuirli. Successivamente vengono esaminate le qualità dei dati connesse alla

conoscibilità e forniti i requisiti principali per il conseguimento di loro adeguati livelli. Infine, vengono formulate raccomandazioni per il legislatore e i decisori istituzionali in materia. Il documento è corredato da una appendice costituita da un glossario nel quale sono date definizioni dettagliate per alcuni termini usati nel testo. I termini per i quali d'ora in avanti viene fornita una definizione, nel testo o nel glossario sono riportati in *grassetto corsivo* la prima volta che compaiono. Il contenuto del presente documento è stato portato all'attenzione dell'Autorità garante per la protezione dei dati personali, che ha fornito un prezioso contributo alla sua stesura.

2. Classificazione dei dati pubblici e loro rappresentazioni

Con il termine *dato* si intende la rappresentazione di un fenomeno osservabile in un *formato codificato*, tale da essere memorizzabile ed elaborabile. Tale formato codificato appartiene in generale a un insieme di possibili *elementi*, chiamato *dominio di definizione* del dato. Il dominio è in genere denotato da un *nome*. Il nome del dominio può essere associato anche al dato. In tale accezione, che assumeremo nel seguito, il dato è una coppia <Nome, elemento di dominio>. Ad esempio il dato <ETÀ, 37>, che rappresenta attraverso un formato codificato (appunto, "37") l'età di una persona fisica, ha come dominio dei possibili elementi l'insieme dei numeri interi da 0 a 150, dominio che si può denotare con il nome ETÀ. Per estensione, ETÀ è anche il nome del dato "37".

L'evoluzione delle tecnologie offre per i dati molte modalità di rappresentazione, oltre quelle tradizionali, quali immagini e suoni.

I dati, sia quando siano memorizzati ed elaborati con tecniche tradizionali, sia, come si assume sempre nel seguito, quando siano rappresentati in modo da essere trattati direttamente con elaboratori digitali, assumono diverse forme:

1. dati strutturati aventi formato ammesso nei campi ed archivi di una base di dati;
2. documenti, ovvero dati non strutturati, caratterizzati dall'essere espressi in linguaggio naturale.

Non interessa distinguere tra tali diverse modalità e neppure tra dati e documenti. In seguito si utilizzeranno, in generale equivalentemente, i termini dato o *insieme di dati* o *dati*, il contesto orienterà ad individuare eventuali differenziazioni. Pur considerando soltanto i dati in formato digitale, molte aspetti del documento possono valere anche per dati rappresentati in formati analogici.

La locuzione *dati pubblici* può essere interpretata secondo diverse accezioni:

1. dati accessibili pubblicamente: questa definizione fa riferimento all'assenza di requisiti di riservatezza, e riflette quindi un aspetto legato alla legittimità della consultazione da parte di soggetti comunque interessati. Risulta quindi applicabile ad un sottoinsieme dei dati detenuti dai soggetti pubblici (ad esempio: le leggi);
2. dati detenuti da un soggetto pubblico: fa riferimento alla natura pubblica del titolare del trattamento (ad esempio: dati anagrafici di persone e imprese) che può eventualmente esserne anche il produttore. È l'accezione adottata dal "Libro verde sull'informazione del settore pubblico nella società dell'informazione" della Commissione europea;
3. dati di interesse di un soggetto pubblico: fa riferimento alla natura pubblica del fruitore nell'interesse della collettività e riguarda informazioni che, anche se detenute da un soggetto privato, devono essere in determinate condizioni rese accessibili ai soggetti pubblici per il perseguimento di fini istituzionali (ad esempio, i dati di tracciamento di contatti telefonici).

Le considerazioni che seguono si applicano al più vasto insieme di dati che risulta dalla unione di quelli individuati con ciascuna delle definizioni sopra riportate.

I dati pubblici possono essere classificati secondo diverse caratteristiche: Quelle che rilevano qui sono:

1. identificabilità;
2. presenza in registri pubblici o simili;
3. aggregazione e generalizzazione;
4. grado di elaborazione;
5. utilità per i soggetti interessati ad accedervi.

Identificabilità

I dati possono essere riferiti a soggetti, persone fisiche o giuridiche, identificati o identificabili, nel qual caso sono **dati personali**, oppure non essere riconducibili a singole persone fisiche o giuridiche, nel qual caso sono **dati anonimi**. Alcuni dati personali, per la delicatezza che li caratterizza, sono definiti **sensibili**, idonei cioè a rivelare l'origine razziale ed etnica, le convinzioni religiose, filosofiche o di altro genere, le opinioni politiche, l'adesione a partiti, sindacati, associazioni od organizzazioni a carattere religioso, filosofico, politico o sindacale, lo stato di salute e la vita sessuale.

Presenza in registri pubblici

In alcuni casi i dati sono **provenienti da pubblici registri, elenchi, atti o documenti conoscibili da chiunque**, senza condizioni. Essi sono tenuti o formati da uno o più soggetti pubblici, in virtù di una norma di legge o di regolamento; la norma che è alla base della conoscibilità in questi casi può prevedere particolari modalità di accesso o limiti temporali che vanno rispettati anche in caso di comunicazione o di diffusione dei dati. Il solo fatto di poter rinvenire un dato personale, ad esempio l'indirizzo di posta elettronica, in uno spazio pubblico di internet, non comporta che esso sia conoscibile da chiunque e quindi non autorizza l'uso libero del dato.

Aggregazione e generalizzazione

I dati sono detti **elementari** quando rappresentano un aspetto della realtà non ulteriormente riconducibile, date le ipotesi adottate, ad aspetti più semplici, **statistici** quando sono il risultato di una elaborazione attraverso funzioni di aggregazione su dati elementari. Un esempio di dato elementare è l'età di una persona; esempi di dato statistico sono il numero di soggetti con quell'età in un territorio specifico e l'età media della popolazione di un comune.

Secondo un diverso processo di astrazione, usualmente detto di generalizzazione, si distinguono i **metadati** e gli **schemi di dati**. I metadati sono proprietà di un insieme di dati. Uno schema di dati consiste nella descrizione di un insieme di classi di dati e delle relazioni che intercorrono tra di essi. Ad esempio:

1. un quadro di un orario ferroviario è un insieme di dati;
2. la proprietà per cui gli orari di arrivo e partenza siano rappresentati mediante due cifre per l'ora e due cifre per il minuto è un metadato;

3. la conoscenza della relazione tra treni e città collegate da treni attraverso gli orari di arrivo e di partenza dei treni in o rispettivamente da tali città è lo schema dei dati del quadro.

I metadati e gli schemi di dati sono in genere anonimi, perché non fanno riferimento ad uno specifico soggetto fisico o giuridico.

Grado di elaborazione

Questa classificazione, tratta dall'importante rapporto Mandelkern¹ commissionato dal governo francese, tiene conto delle elaborazioni cui i dati sono stati sottoposti. Secondo tale classificazione, i dati pubblici possono essere così suddivisi:

a. **dati grezzi**: sono i dati raccolti ma non ancora sottoposti ad elaborazioni significative e che quindi si trovano sostanzialmente nella forma in cui sono stati acquisiti, quali:

1. dati acquisiti tramite digitazione da moduli di carta (oppure via internet tramite moduli elettronici), prima delle verifiche tese a confrontarne il contenuto con altri dati contenuti in archivi di riferimento o raccolti contestualmente;
2. messaggi ricevuti per posta elettronica prima di operazioni di marcatura del testo le quali ne caratterizzano specifiche parti o informazioni presenti;
3. dati geografici prima delle operazioni di normalizzazione;

b. **dati di base**: sono i dati già sottoposti alle elaborazioni necessarie a renderli elaborabili al di fuori di un singolo sistema o di una singola tecnologia, solitamente da parte di soggetti diversi da quello che li ha raccolti, ad esempio:

1. indirizzi postali normalizzati;
2. nominativi di professionisti iscritti ad un albo;

c. **dati arricchiti** (o **elaborati**): sono i dati risultanti da operazioni di ricerca e di confronto con informazioni di differente provenienza, ma collegate ad uno stesso fenomeno; la categoria contiene anche i dati aggregati in senso statistico, come medie, indici ed altro. Esempi di dati arricchiti sono:

1. la posizione fiscale di un'impresa come risulta da diverse basi di dati del Ministero dell'economia e finanze;
2. i dati contenuti nello stato di famiglia di un cittadino;
3. l'indice dei prezzi al consumo collegato ad una città;
4. una carta geografica ricavata da dati fotogrammetrici grezzi.

Alle categorie di dati caratterizzate nelle due precedenti classificazioni (aggregazione e grado di elaborazione) si applicano le considerazioni che seguono.

Le grandi basi di dati gestite da soggetti pubblici, spesso a copertura nazionale, rendono sempre più facile produrre, attraverso semplici elaborazioni, nuovi dati a partire da quelli già disponibili. Tali operazioni possono essere compiute per fini istituzionali all'interno di uno stesso soggetto, oppure sfruttando almeno in parte dati presenti presso altri soggetti.

I dati di base, essendo solitamente normalizzati o standardizzati, possono essere messi in relazione con dati di diversa provenienza e, in particolare, essere collegati a persone fisiche o giuridiche. In

¹ D. Mandelkern, *Diffusion des données publiques et révolution numérique* Commissariat général du Plan, 1999 (reperibile su <http://www.ladocfrancaise.gouv.fr>).

un periodo di rapida evoluzione tecnologica è problematico definire quali dati siano personali, in quanto un soggetto può essere identificato anche tramite l'elaborazione e l'incrocio di dati che - presi singolarmente - sono anonimi. Si introduce, perciò, il concetto di *mezzi ragionevoli* secondo il quale un soggetto può essere ritenuto identificabile quando, con l'impiego di mezzi ragionevoli (tempo, risorse, eccetera) sia possibile stabilire una relazione tra dati apparentemente anonimi e dati identificativi del soggetto stesso. La valutazione di mezzi ragionevoli non può, peraltro, essere disgiunta dal grado di interesse economico o di altra natura a conoscere dati personali.

La disponibilità di archivi conoscibili via internet permette altresì di ricavare nuovi dati i quali presentano un grado di riservatezza maggiore rispetto ai dati di partenza. Diviene quindi necessaria l'identificazione di linee guida tecniche che impediscano, o per lo meno rendano non possibile con mezzi ragionevoli, la costruzione di archivi derivati i cui dati prevedano requisiti diversi per la conoscibilità rispetto a quelli imposti per i dati originari utilizzati.

Utilità per i soggetti interessati ad accedere al dato

I *dati essenziali* sono i dati pubblici dei quali i cittadini, le imprese e altri operatori privati devono poter disporre per esercitare i propri diritti. I dati essenziali possono essere sia anonimi, come le norme e molti dati statistici, sia personali (la maggior parte dei dati amministrativi); per questi ultimi è necessario mettere a punto regole che ne assicurino la protezione senza ostacolarne la conoscibilità da parte degli aventi diritto. Pur non esistendo una definizione giuridicamente rilevante dei dati essenziali, si possono fare rientrare in questa categoria i dati la cui conoscibilità è sancita da leggi, ad esempio quella relativa alla trasparenza amministrativa. La seguente è una lista, indicativa e non esaustiva, di dati da considerare essenziali:

1. le leggi e i regolamenti vigenti;
2. i dati statistici nazionali più importanti o necessari per le decisioni individuali o collettive;
3. i dati personali in possesso dei soggetti pubblici e che riguardano il richiedente;
4. le indicazioni necessarie ad usufruire dei servizi erogati da soggetti pubblici e a verificare lo stato dell'iter amministrativo (portali e sportelli di accesso unificati ai servizi, organigrammi, indirizzi postali, numeri di telefono, eccetera).

Secondo un'interpretazione estensiva, si potrebbe includere tra i dati essenziali per un cittadino, o un'impresa o un altro operatore privato, tutti i dati ad essi relativi in possesso di soggetti pubblici: ne consegue che potrebbe essere garantito non soltanto il diritto di conoscerli, ma anche quello di aggiornarli, sia pure secondo procedure individuate da ciascun soggetto pubblico o concordate fra più soggetti pubblici, comunque rese note e, se necessario, sancite da norme.

Sempre estensivamente, potrebbero essere considerati essenziali per un soggetto pubblico i dati, di qualsiasi natura, in possesso di altri soggetti pubblici e necessari al primo per adempiere ai propri compiti istituzionali.

In generale, per quanto riguarda la conoscibilità dei dati pubblici si ritiene prioritaria l'identificazione dei dati che devono essere considerati essenziali: un patrimonio da mettere a disposizione di ogni cittadino, con la consapevolezza che le tecnologie permettono ormai di farlo in forma digitale e a costi trascurabili.

Come emerge dalla trattazione effettuata, le precedenti tipologie di dati pubblici presentano diverse aree di sovrapposizione; ad esempio esistono dati statistici anonimi e dati statistici personali, ovvero i dati sensibili sono una sottospecie dei dati personali.

3. Modalità di attuazione della conoscibilità e soggetti interessati

La conoscibilità dei dati pubblici può essere consentita attraverso tre modalità di scambio, diverse a seconda dei ruoli, attivi o passivi, che assumono l'amministrazione che detiene i dati e il soggetto che intende conoscerli:

1. l'accesso è di iniziativa del soggetto interessato e gli permette di esprimere le sue esigenze; gli accessi possono essere effettuati sia da persone sia da sistemi informatici;
2. la comunicazione è di iniziativa dell'amministrazione oppure dell'interessato che ne ha fatto richiesta e consiste nel far pervenire i dati ad uno o più destinatari predeterminati; può essere unidirezionale o bidirezionale; può essere preceduta da una richiesta o meno;
3. la diffusione è di iniziativa della amministrazione e consiste nel rendere i dati disponibili ad una platea indeterminata di soggetti anche tramite pubblicazione, tradizionale o su internet; il grado di effettiva disponibilità dipende dalle modalità prescelte.

I dati pubblici possono essere resi disponibili da **distributori pubblici**, in generale coloro che li producono o li detengono, solitamente nella forma di dati di base, e **distributori privati**.

I soggetti fruitori possono essere:

1. l'insieme delle pubbliche amministrazioni;
2. particolari pubbliche amministrazioni;
3. soggetti privati, fisici o giuridici;
4. categorie professionali, come avvocati, notai, commercialisti, eccetera.

Si possono avere quattro situazioni distinte per quanto concerne la conoscibilità dei dati detenuti da soggetti pubblici:

1. dati **conoscibili da chiunque**;
2. dati **a conoscibilità circoscritta** riservata ad alcuni soggetti pubblici, oppure ad alcune categorie professionali o ad altre particolari categorie di soggetti;
3. dati **conoscibili** ai sensi dell'articolo 22 della legge n. 241/1990 da parte di chi ha un interesse personale e concreto e per la tutela di situazioni giuridicamente rilevanti;
4. dati **conoscibili dal solo soggetto pubblico** che li detiene, coperti cioè dal segreto d'ufficio o dal segreto statistico o sottoposti a particolare tutela dalla legge n. 675/99.

4. Qualità dei dati utili per la conoscibilità

Una **qualità dei dati** è una caratteristica desiderata da parte degli utenti. Legate alla conoscibilità sono due qualità: sicurezza e usabilità. Inoltre, è importante che i dati siano aggiornati, che siano esatti, cioè corrispondano in modo corretto (se ciò sia possibile) al fenomeno osservato, ovvero almeno accurati, cioè ne siano una buona approssimazione.

Sicurezza

Gli aspetti di sicurezza coinvolgono tutti i settori dell'informatica, sotto il profilo della sicurezza sia logica, intesa come insieme di caratteristiche e procedure per la salvaguardia dei dati, sia fisica, intesa come insieme di caratteristiche e procedure per la salvaguardia delle infrastrutture utilizzate. Ogni soggetto pubblico deve, quindi, prendere misure atte a:

1. garantire la **disponibilità** delle risorse, compresi i dati allorché conoscibili, e dei servizi del sistema informatico;
2. impedire attacchi rivolti a violare la **riservatezza** dei dati e delle informazioni, consentendone la fruizione soltanto a persone o sistemi informatici autorizzati;
3. assicurare l'**integrità** dei dati e delle informazioni e più in generale delle risorse, non consentendo modifiche non autorizzate.

Per una trattazione più ampia dei diversi aspetti della sicurezza e delle azioni necessarie per garantirla, anche con riferimento alla sicurezza dei dati, si vedano le “Linee guida sulla sicurezza”, approvate dall'Aipa il 15 novembre 2001 e pubblicate sul sito www.aipa.it. In termini generali, si può affermare che le misure di sicurezza vanno graduate in funzione delle informazioni trattate e del rischio che, per la loro natura, siano soggette ad intrusioni, minacce o attacchi (esterni o interni). Per fronteggiarli, si apprestano contromisure che possano ridurre i rischi di danneggiamento alle risorse, con un ragionevole bilanciamento tra i costi da sostenere e la riduzione del livello di rischio.

Usabilità

In questo contesto intendiamo per usabilità la facilità con cui un soggetto interessato a conoscere il dato riesce ad accedervi con le tecnologie disponibili e a partire dalla propria situazione fisica, psichica e culturale; è quindi una qualità che dipende sia dalle tecnologie informatiche che possono essere usate per l'accesso sia dalle caratteristiche dell'utente. Ad esempio, per un utente che abbia un vocabolario limitato l'usabilità può essere migliorata dotando l'interfaccia di accesso di icone e percorsi che orientino intuitivamente sul modo di formulare la richiesta. Per un non vedente, accedere ad un dato disponibile su un sito internet è impossibile in assenza di tecnologie assistive quali *screen reader*; per un utente che conosca soltanto la lingua italiana risulterà di scarso utilizzo un sito in lingua inglese. In particolare, per un utente non abituato ad utilizzare linguaggi di interrogazione, una base dati può essere resa meglio accessibile da un'interfaccia che lo guidi nella formulazione delle richieste.

L'usabilità può essere dettagliata nelle seguenti caratteristiche:

1. **apprendibilità**, intesa come l'impegno richiesto agli utenti per comprendere le possibilità d'impiego delle tecnologie necessarie per accedere ai dati;
2. **comprensibilità**, intesa come l'impegno richiesto agli utenti per capire il funzionamento delle tecnologie;
3. **operabilità**, intesa come l'impegno richiesto agli utenti per accedere ai dati attraverso le tecnologie.

I tre aspetti possono essere misurati rispettivamente attraverso il tempo necessario per l'apprendimento, il tempo necessario per acquisire l'abilità necessaria all'utilizzo e il tempo richiesto per l'effettivo accesso ai dati.

Altre qualità dei dati

Le principali ulteriori caratteristiche di qualità dei dati fanno riferimento a:

1. **esattezza**, la qualità per cui i dati corrispondono correttamente, se ciò sia possibile, alle caratteristiche osservate dal fenomeno di interesse; poiché l'esattezza può essere ottenuta in generale soltanto con alti costi, in diverse circostanze ci si accontenta della
2. **accuratezza** ed in particolare:

- a. **accuratezza sintattica** è la vicinanza del dato ad un elemento del dominio di definizione. Ad esempio, il dato <NOME, Mrio> è inaccurato sintatticamente perché diverso dal dato <NOME, Mario>, appartenente al dominio dei nomi di persona; l'approssimazione può essere tuttavia sufficiente a consentire la necessaria correzione;
- b. **accuratezza del significato** è la vicinanza del dato all'elemento del dominio di definizione che corrisponde al fenomeno osservato. Ad esempio, il dato <NOME, Mario> è inaccurato (rispetto al significato) se la persona cui fa riferimento si chiama "Maria"; la conoscenza di altre caratteristiche può consentire la necessaria correzione;
3. **completezza** (di un insieme di dati), cioè la copertura con la quale l'insieme osservato è rappresentato nell'insieme di dati. Ad esempio, per gli impiegati di un'organizzazione, un'anagrafica è completa quando sono rappresentati tutti gli impiegati;
4. **consistenza interna** (di un insieme di dati) consiste nel rispetto dei vincoli di significato che legano logicamente l'insieme dei dati. Ad esempio, la coppia di dati <CAP, 35122>, <CITTA', Roma>, che rappresenta codice di avviamento postale e città di un indirizzo toponomastico, non è caratterizzata da consistenza interna perché non è rispettato il legame funzionale tra il codice e la città;
5. **consistenza esterna** (dello stesso dato in diversi insiemi di dati) è la proprietà per cui il dato è rappresentato con lo stesso elemento nei diversi insiemi. La consistenza esterna, così come definita, assume che il vincolo di significato sia l'identità. La consistenza esterna può essere persa nei processi di trasformazione che il dato subisce, dando luogo alla produzione di copie dello stesso dato non coincidenti;
6. **tempestività**, per cui un mutamento del fenomeno rappresentato porta all'aggiornamento del dato in tempo utile rispetto alla sua utilizzazione;
7. **pertinenza e non eccedenza**, rapportate al fine per il quale venga posto in essere un trattamento e collegate al rapporto tra dati e adempimenti.

Il problema del mantenimento di un'elevata qualità dei dati è solitamente affrontato identificando:

1. i requisiti di qualità dei dati che costituiscono il patrimonio informativo di un soggetto;
2. gli indicatori di qualità, coinvolgendo sia il responsabile del mantenimento dei dati sia gli utilizzatori, unitamente alle modalità di valutazione e in relazione agli obiettivi da conseguire;
3. i flussi di dati, sia interni sia provenienti dall'esterno, i corrispondenti livelli di qualità e i responsabili del loro mantenimento.

È opportuno collegare esplicitamente i concetti di origine informatica alle previsioni normative: ad esempio, il titolare del trattamento dei dati può essere distinto dalla persona che deve curarne la qualità, indicato nell'ambiente informatico con i termini **data steward** o **responsabile della qualità dei dati**; ciò permetterebbe di precisare meglio gli obblighi di ciascuno.

Problemi possono sorgere quando esistano diverse rappresentazioni di uno stesso dato o documento. Anche quando esse vengano diffuse dallo stesso soggetto titolare e sia quindi ragionevole ritenere che possiedano la stessa attendibilità, può non essere così. Soggetti pubblici hanno fatto ricorso recentemente alla diffusione in rete di versioni elettroniche di documenti, precisando di riconoscere validità legale soltanto alla pubblicazione a stampa o a specifiche rappresentazioni informatiche dei dati; in altri casi, viceversa, il ricorso alle tecnologie informatiche ha di fatto esteso il concetto di copia conforme di un documento, potendo per questa via innovare l'interazione tra cittadino e soggetti pubblici.

Le modalità di raccolta, elaborazione ed arricchimento e le scelte tecnologiche hanno conseguenze sulla qualità dei dati forniti ai cittadini, quindi sulla qualità dei servizi resi. Le principali caratteristiche di qualità dipendono direttamente dal grado di controllo che il soggetto detentore dei dati riesce ad esercitare sulle diverse fasi. Nonostante l'impiego di tecnologie avanzate, si osserva spesso un'insufficiente qualità dei dati acquisiti dai cittadini e dalle imprese. I soggetti pubblici tendono a conservare il ruolo di collettori ma raramente di certificatori dei dati: la conseguenza più frequente è la ripetizione periodica dell'acquisizione e talvolta i dati già posseduti sono di qualità migliore di quella che ci si può attendere dai nuovi.

La conoscibilità da parte dei diretti interessati dei dati già in possesso dei soggetti pubblici, oltre che essere strumento di trasparenza, può consentire una maggiore efficienza. Bisogna anche considerare che le scelte organizzative hanno un impatto sulla qualità dei dati maggiore di quelle tecnologiche: ad esempio, l'acquisizione di dati attraverso un servizio online che permetta di effettuare comunicazioni formali ad un soggetto pubblico può portare ad un notevole miglioramento nell'accuratezza.

Un altro aspetto da tenere presente riguardo all'accesso è che esso non deve creare discriminazioni per particolari categorie di utenti: una norma pubblicata sulla Gazzetta ufficiale cartacea è accessibile più facilmente in Italia che all'estero perché c'è un maggior numero di punti di vendita e di consultazione; la stessa norma pubblicata su internet è consultabile da chiunque, indipendentemente dalla residenza, abbia a disposizione un elaboratore, un collegamento ad internet ed il software adatto per leggere il documento.

5. Requisiti di qualità

Un soggetto pubblico che intenda rispettare i profili di conoscibilità definiti in precedenza deve assicurare elevati livelli di sicurezza e di usabilità. Inoltre, deve assicurare ai dati un elevato livello complessivo di qualità. Nelle prossime tre sezioni sono trattati distintamente gli aspetti relativi a sicurezza, usabilità ed altre qualità, definendo corrispondenti criteri e linee guida.

5.1 Requisiti per la sicurezza

La quantità di dati pubblici disponibili e la diffusione delle tecnologie informatiche e di internet pongono rilevanti problemi di sicurezza ai soggetti che detengono i dati, riguardo sia alla gestione dei dati sia alla loro conoscibilità. La individuazione delle misure di sicurezza più opportune da adottare all'interno del dominio organizzativo di un soggetto pubblico non costituisce oggetto di questo documento e si rimanda alle "Linee guida sulla sicurezza" già richiamate in precedenza. Le misure di sicurezza da adottare nei confronti di fruitori esterni sono, invece, descritte nel seguito.

Poiché la maggior parte dei soggetti pubblici è dotata di siti internet tramite i quali è possibile accedere ai dati prodotti, una misura minima di sicurezza per la protezione da attacchi esterni consiste nella separazione fisica dei server che contengono i dati pubblici oggetto di possibile accesso dai sistemi informatici interni che contengono l'insieme dei dati pubblici in possesso di ciascun soggetto. Ci si può proteggere così contro il vandalismo informatico e contro il rischio di accessi indiscriminati a dati non conoscibili da chiunque.

Procedure di sicurezza nel caso di accesso da parte dell'utente

La conoscibilità dei dati mediante l'accesso può essere permessa soltanto da parte dei soggetti pubblici in grado di assicurare la qualità dei dati ed eventualmente ai soli dati che non vincolano il soggetto pubblico nel caso di loro utilizzo da parte di terzi.

I requisiti richiesti per l'accesso e i suoi limiti variano a seconda della tipologia di dati a cui accedere e delle caratteristiche dell'utente (persona o sistema informatico).

Nel caso in cui il fruitore sia una persona fisica, i processi da svolgere sono:

1. **identificazione**, che può essere effettuata attraverso l'assegnazione di un identificativo utente;
2. **autenticazione**, che - a seconda del livello di sicurezza richiesto - può consistere nella assegnazione di una *password* o nell'utilizzo di un sistema a chiavi pubbliche (PKI - *Public Key Infrastructure*);
3. **autorizzazione**, che prevede l'assegnazione di un *profilo di accesso* all'utente (per esempio autorizzazione alla sola lettura, alla lettura e all'aggiornamento, alla cancellazione, alla stampa, eccetera) e il controllo della congruità della tipologia di richiesta ad ogni accesso.

I processi descritti possono essere resi più complessi in funzione del livello di sicurezza richiesto dalle caratteristiche delle informazioni: se, ad esempio, si richieda un livello di sicurezza elevato, si possono prevedere strumenti di registrazione degli accessi per tenere traccia di quando siano avvenuti e da parte di chi, oppure si può richiedere all'utente la fornitura di una impronta biometrica o di una firma digitale su un registro di sicurezza informatico.

Nel caso in cui l'accesso alle informazioni avvenga da parte di un sistema informatico e non di una persona, le funzioni di sicurezza da prevedere sono le stesse, ma potrebbero essere realizzate con modalità diverse: in questo caso non è possibile fornire informazioni di natura personale o biometrica, ma possono essere richieste informazioni che caratterizzino puntualmente il sistema richiedente, come per esempio l'*IP address* del richiedente, o può essere utilizzato un sistema a chiave pubblica, metodo previsto da tecniche quali il Kerberos o IPSEC.

Se fosse necessario garantire la riservatezza dei dati durante il processo di trasporto telematico sino all'utente, il soggetto pubblico che detiene i dati potrebbe utilizzare tecniche di crittografia e protocolli quali SSL o HTTPS, o mettere a disposizione del trasporto canali ad elevata sicurezza.

Un ultimo aspetto da considerare nel caso di accessi telematici è il **non ripudio** di aver reso accessibile il dato da parte di chi lo detiene, aspetto che assume particolare rilevanza in relazione alla tipologia di dati e ai profili di responsabilità. Nella tabella che segue sono riassunti i requisiti per l'accesso telematico, in relazione alle diverse tipologie di dati pubblici. Il non ripudio da parte di chi detiene i dati è relativo all'averli resi disponibili.

	Requisiti per l'accesso			
	Fruitore (utente)			Detentore
Tipologie di dati	Identificazione	Autenticazione	Autorizzazione	Non ripudio
conoscibili da chiunque	no	no	no	no
a conoscibilità circoscritta ad alcuni soggetti pubblici	sì	sì	no	no
a conoscibilità circoscritta ad alcune categorie professionali	sì	sì	sì	no
accessibili ai sensi della legge n.241/90	sì	sì	no	sì
Accessibili al solo soggetto pubblico che li detiene	sì	sì	sì	sì
Sensibili	sì	sì	sì	sì

Qualora l'accesso ai dati esposti dal soggetto pubblico avvenga tramite un sistema informatico che proceda autonomamente al prelievo dei dati di interesse, si presenta un ulteriore problema dovuto al fatto che il sistema informatico del soggetto che detiene i dati può essere momentaneamente indisponibile, o momentaneamente soggetto ad aggiornamento, ovvero che la richiesta riguardi un volume di dati tale da rendere il sistema indisponibile ad altri accessi per molto tempo. Una soluzione adottata per fronteggiare il primo inconveniente è quella di separare gli archivi operativi del soggetto pubblico da quelli ove risiedono i dati esposti; per il secondo aspetto è necessario verificare il livello di tempestività connesso alla richiesta; per il terzo aspetto può essere opportuno mantenere diverse copie degli archivi. In ogni caso, insorge il problema dell'utilizzabilità dei dati ai quali potrebbe non essere attribuito valore giuridico in assenza di una firma che ne attesti la corrispondenza con quelli detenuti dal soggetto pubblico.

Requisiti di sicurezza nel caso di comunicazione e diffusione da parte del soggetto titolare

La comunicazione telematica di dati pubblici da parte del soggetto detentore ad un utente, sia esso una persona o un sistema informatico, può avvenire di iniziativa del detentore al verificarsi di un evento che riguardi l'utente, oppure su iniziativa e richiesta dell'utente.

Nel primo caso, è necessaria la verifica da parte del soggetto pubblico che effettivamente l'evento interessi l'utente e che questi abbia diritto ad esserne informato (processo di autorizzazione). Nel secondo caso, la comunicazione viene attivata dall'utente e quindi, relativamente alla richiesta, continua a valere quanto già descritto in precedenza.

Nel caso della comunicazione il *non ripudio* assume un significato più ampio, caratterizzando l'impossibilità per chi ha ricevuto i dati di negarne la ricezione e, per chi li ha forniti, di negarne la comunicazione.

Anche nei casi di comunicazione e diffusione, è il soggetto pubblico che, se necessario, è tenuto ad assicurare la riservatezza e l'integrità dei dati trasmessi. In generale:

1. se le informazioni devono essere trasmesse in maniera tale che il solo destinatario possa conoscerle, è necessario ricorrere a tecniche di crittografia, per assicurarne la riservatezza;
2. se è necessario garantire aspetti di integrità alle informazioni, devono essere previste tecniche in grado di assicurare la non modificabilità delle informazioni durante la trasmissione;
3. se è necessario assicurare da parte del mittente requisiti di autenticazione, utili anche per il non ripudio, una possibile soluzione è la firma digitale;
4. se è necessario assicurare al mittente il non ripudio da parte del destinatario, occorre prevedere meccanismi di notifica di ricezione.

Nella tabella che segue, dove **M** indica il soggetto mittente dei dati e **D** il soggetto destinatario, sono riassunti i requisiti per la comunicazione telematica dei dati pubblici, in relazione alle diverse tipologie.

	Requisiti per la comunicazione			
	M/D	M	M	D
Tipologie di dati	Non ripudio	Integrità	Riservatezza (crittografia)	Autenticazione
Conoscibili da chiunque	no	no	no	no
a conoscibilità circoscritta ad alcuni soggetti pubblici	sì	sì	no	no
a conoscibilità circoscritta ad alcune categorie professionali	sì	no	no	no
Accessibili ai sensi della legge n.241/90	sì	sì	no	no
accessibili soltanto al soggetto pubblico che li detiene	sì	sì	sì	sì
sensibili	sì	sì	sì	sì

In conclusione il non ripudio da parte del soggetto mittente dei dati è relativo alla trasmissione e, se preceduta da una richiesta, alla ricezione della stessa; il non ripudio da parte del soggetto destinatario è relativo alla ricezione dei dati.

Soprattutto nel caso di comunicazione tra soggetti pubblici, accade spesso che venga contestualmente inviato anche il formato cartaceo ed allora il destinatario deve attenderne l'arrivo e controllare la corrispondenza tra i contenuti; ciò non giova alla tempestività. La limitazione è oramai superabile, in quanto la validità legale dei dati trasmessi può essere garantita, ad esempio, trasmettendo dati sui quali sia stata apposta la firma digitale.

Un ultimo aspetto collocabile nella tematica della sicurezza è quello della utilizzazione a fini pubblici di dati anche sensibili. Esistono molti fini (ad esempio epidemiologici) per cui diverse categorie di dati sensibili possono trovare un'importante valorizzazione quando utilizzati da specifici soggetti. In questi casi occorre definire regole rigorose che tuttavia non devono impedire, ma anzi disciplinare e perciò favorire, l'utilizzazione.

5.2 Requisiti per l'usabilità

L'insieme dei *mezzi di diffusione e accesso* a dati sta diventando sempre più ampio; tra essi possono essere citati siti web, CD Rom, terminali con interfaccia applicativa di accesso ad un sistema informativo. Più in generale, l'usabilità, e le regole e gli strumenti per migliorarla, dipendono fortemente dalla tecnologia utilizzata e dalla tipologia di utenza. È opportuno distinguere due categorie di utenti: soggetti pubblici e soggetti fisici e giuridici costituenti la collettività.

Usabilità per soggetti pubblici

La legge 24 novembre 2000, n. 340 stabilisce che “Le pubbliche amministrazioni di cui all’articolo 1, comma 2 del decreto legislativo n. 29 del 3 febbraio 1993 hanno accesso gratuito ai dati contenuti in pubblici registri, elenchi, atti o documenti da chiunque conoscibili”. Tale norma pone le premesse per una effettiva possibilità delle pubbliche amministrazioni di scambiare dati di proprio interesse utilizzando le tecnologie ICT.

Per realizzare un’interazione automatica tra sistemi gestiti da organizzazioni distinte è indispensabile definire di protocolli comuni attraverso cui attuare il colloquio. La generale condivisione dei protocolli di comunicazione utilizzati dalla rete internet è stata determinante per l’estensione mondiale della sua diffusione e della sua continua crescita.

Le regole da seguire sono necessariamente differenti per basi di dati strutturati e basi documentali: infatti per i dati strutturati sono previste tecnologie, modelli e linguaggi (sistemi di gestione di basi di dati) che permettono di accedere e manipolare in modo preciso la conoscenza rappresentata; i documenti testuali possiedono un contenuto informativo più ricco, ma i relativi sistemi di gestione non presentano attualmente modelli di rappresentazione formalizzati per i quali sia possibile realizzare funzionalità analoghe.

Esistono poi due livelli tecnologici ai quali si può fare riferimento nel dare indicazioni alle amministrazioni: il livello del **dominio dell’amministrazione**, cioè delle risorse tecnologiche gestite autonomamente dalla amministrazione e il livello del **sistema distribuito cooperativo** al quale l’amministrazione partecipi. All’interno del sistema distribuito cooperativo possiamo distinguere i **sistemi di interfaccia** tra il dominio della amministrazione e il sistema distribuito, detti anche **porte applicative**, e il **sistema di gestione della cooperazione**. Valgono i seguenti criteri.

a. Dominio dell’amministrazione

1. per i dati strutturati, è opportuno adottare il modello relazionale e sistemi di gestione relazionali, che rappresentano ormai uno standard universalmente accettato;
2. per le basi documentali, seppure non sia possibile individuare un modello di rappresentazione comune a tutte le tipologie di documenti, è opportuno associare ad esse una struttura, attraverso l’utilizzo di un linguaggio di marcatura come l’XML, ormai affermatosi come standard *de facto*.

b. Sistema distribuito cooperativo

1. i dati strutturati devono essere rappresentati nei sistemi di interfaccia mediante modelli concettuali che permettano di evidenziarne il significato più che gli aspetti realizzativi. Per tali modelli esiste una varietà di proposte (ad esempio, modello entità-relazione, UML, XML) ma non uno standard affermato. Soltanto in alcuni casi (XML) esistono linguaggi di interrogazione che rendono la rappresentazione concettuale operativamente utilizzabile per l’accesso;
2. riguardo alle basi documentali, è opportuno che a ciascun documento sia associato nel sistema di interfaccia un nome uniforme, per identificare univocamente il documento nel sistema distribuito, indipendentemente dalla localizza fisica, adottando regole analoghe a quelle espresse per i documenti giuridici nella circolare Aipa 35/2001;
3. il sistema di gestione della cooperazione deve catalogare in modo integrato le informazioni relative ai dati strutturati e alle basi documentali disponibili sui sistemi di interfaccia, per quanto riguarda i contenuti, le modalità e le restrizioni di accesso e quanto altro necessario a consentirne l’utilizzo. Tale catalogo costituisce un “registro certificato” che funge da indice e indirizzario dei servizi che ciascun dominio offre ai potenziali utilizzatori;

4. tutte le volte che le amministrazioni scambiano dati strutturati e documenti attraverso il sistema distribuito cooperativo, è opportuno che venga adottato un formato di scambio basato sul linguaggio XML, adottando per i dati più significativi o di più frequente utilizzo un formato comune, il quale deve essere definito dal sistema di gestione della cooperazione. Questo criterio è stato adottato, relativamente ai dati scambiati tramite posta elettronica, nelle regole tecniche DPR 445 del 28 dicembre 2000, che hanno definito un formato obbligatorio per la segnatura dei messaggi e fornito indicazioni per un insieme più ampio di tipologie di dati.

Usabilità per la collettività

Quanto più l'utente si discosta, nelle sue caratteristiche fisiche, psichiche e culturali, da una generica condizione di "normalità", tanto più la tecnologia informatica dovrà prevedere strumenti che elevino il grado di usabilità. Possiamo individuare tre tipologie di utenti, caratterizzate a loro volta da molte sfumature intermedie:

1. **utente "normale"**, caratterizzato da un grado di cultura complessiva e da una conoscenza di base degli strumenti informatici tali da far ritenere che lo sforzo necessario per aumentare la usabilità, sul versante sia delle tecnologie sia della formazione e addestramento, sia limitato;
2. **utente con divario culturale**, caratterizzato da limitazioni nella cultura, nel linguaggio, nella capacità di utilizzo delle tecnologie dell'informazione, che impedirebbero l'accesso a molte tipologie di informazioni rese fruibili a soggetti "normali" e che rendono perciò necessario l'utilizzo di tecnologie ad alto grado di usabilità per permettere l'effettivo accesso;
3. **utente disabile**, caratterizzato da limitazioni fisiche o psichiche che rendono indispensabile l'utilizzo di specifiche tecnologie assistive le quali gli permettono di accedere al dato.

Le tecnologie informatiche che rendono possibile l'accesso ai dati riguardano almeno tre strati diversi:

1. **l'interfaccia utente**, attraverso cui l'utente e il sistema informatico interagiscono per la formulazione delle esigenze e la restituzione dei risultati;
2. il **programma applicativo**, cioè l'insieme delle funzionalità che sono messe a disposizione dell'utente per soddisfare le sue esigenze;
3. il **contenuto**, cioè l'insieme dei dati che possono essere resi disponibili, nel formato in cui sono rappresentati.

L'incrocio fra tipologie di utenti e tecnologie porta a definire strategie diverse per ciascuna delle corrispondenze. Le problematiche di usabilità legate all'interfaccia, agli applicativi e agli aspetti di contenuto per gli utenti disabili sono state affrontate in ambito italiano dalla circolare n. 3/2001 del 13 marzo 2001 del Dipartimento della funzione pubblica e dalla circolare dell'Autorità per l'informatica nella pubblica amministrazione, n. 32 del 6 settembre 2001 alle quali si rimanda. In quest'ultima circolare è stato utilizzato il termine **accessibilità** con significato sostanzialmente analogo all'accezione qui adottata di usabilità.

L'usabilità può essere garantita rispettando i seguenti principi e criteri generali:

1. le specifiche di progetto devono sempre tener conto della varietà di esigenze di tutti gli utenti (principio della progettazione universale, *design for all*). Il principio è nato in ambito architettonico; un esempio di applicazione è la previsione di rampe in luogo o in aggiunta alle scale;

2. vi deve essere una chiara separazione tra interfaccia e programma applicativo e tra programma applicativo e contenuto, in modo che sia possibile adattare la presentazione a particolari tipologie di utenti;
3. la progettazione dell'interfaccia, delle funzionalità applicative e del contenuto deve avvenire tenendo presenti le esigenze di tutti gli utenti; esse devono essere inizialmente raccolte direttamente dagli utenti e devono essere verificate su un primo prototipo. Si possono ottenere così notevoli economie rispetto all'approccio in cui l'adattamento alle esigenze degli utenti venga effettuato in fase di manutenzione evolutiva;
4. occorre erogare servizi di *call center* per fornire ai fruitori che la richiedano un'assistenza tecnica veloce e di elevata qualità;
5. vanno garantiti una finestra temporale di erogazione del servizio di accesso e un predefinito livello di qualità del servizio, in termini almeno di tempi di disponibilità e di risposta.

L'ultimo criterio è importante nel caso di accesso ai dati e meno stringente per comunicazione e diffusione.

Possono essere introdotti altri criteri specifici, fra i quali :

per l'interfaccia utente

- è opportuno prevedere l'utilizzo di comandi semplici da interpretare, rappresentati mediante icone o linguaggi che siano universalmente comprensibili;
- quando sia richiesto l'apprendimento di un linguaggio tecnico, occorre rendere disponibili *tour* guidati, dimostrazioni, esempi che aiutino l'utente a familiarizzare con l'interfaccia;
- l'interfaccia deve essere consistente al suo interno, evitando di usare linguaggi e messaggi tra loro contraddittori o non coordinati;

per le funzionalità applicative

- deve essere limitato e di univoca interpretazione l'insieme delle funzionalità disponibili all'utente per esprimere l'esigenza (interrogazione), evitando eccessivi livelli di menù, navigazioni destrutturate, fasi intermedie in cui l'utente possa disorientarsi;
- deve essere usato un opportuno insieme di funzionalità basate su diversi paradigmi di accesso: interrogazione (*query*), *information retrieval*, navigazione ipertestuale, ricerca in linguaggio naturale, accesso proattivo (l'utente è informato periodicamente su particolari tipologie di dati per i quali abbia, ad esempio, sottoscritto un abbonamento), accesso guidato, eccetera;
- deve essere mantenuto, se possibile, un profilo utente che permetta ad ogni nuova interazione di adattare le funzionalità (e l'interfaccia) alla particolare tipologia di utilizzo;

per il contenuto

- si deve prevedere un'organizzazione dei dati basata su archivi omogenei e, se utile, correlati tra di loro, in modo tale da facilitare la comprensione e l'eventuale elaborazione da parte dell'utente;
- si devono usare rappresentazioni con formati immediatamente comprensibili, con l'utilizzo di tabelle, diagrammi, grafici, icone, linguaggio naturale, che permettano la massima usabilità;
- occorre evitare la coesistenza di versioni parallele che rende difficoltosi gli aggiornamenti.

5.3 Requisiti per le altre qualità

Alcuni ulteriori requisiti di qualità per i dati resi conoscibili sono i seguenti:

1. devono essere definite procedure organizzative e strumentali per la misura (e il miglioramento) di tutte le qualità definite in precedenza e cioè esattezza, accuratezza, completezza, consistenza, tempestività, pertinenza e non eccedenza;
2. deve essere garantito l'aggiornamento dei dati, con impegni predefiniti di frequenza e tempestività;
3. deve essere garantita la rappresentazione storica dei dati per un determinato periodo;
4. nel caso di dati dai quali possano derivare effetti rilevanti, si deve indicare in maniera non equivocabile se si tratti di dati ancora soggetti a verifica e la data del loro ultimo aggiornamento, in particolar modo se la comunicazione è stata attivata da una richiesta dell'interessato.

Inoltre, se un dato è errato e viene riconosciuto come tale la prima volta che viene utilizzato, è spesso possibile, anche se non necessariamente agevole, correggere l'errore prima che i danni provocati divengano ingenti. Spesso possibili errori vengono individuati molto tempo dopo che sono stati generati e soltanto indirettamente, in base ad altri errori causati da quelli originari: in questo caso la prima difficoltà consiste nel risalire alla causa che ha dato luogo al primo errore della catena, che può ovviamente essere banale, per poi correggere tutti gli errori conseguenti. È inoltre evidente che se il dato è stato comunicato o diffuso, all'interno e/o all'esterno del soggetto presso il quale si è verificato l'errore, vi sono alte probabilità che si siano verificati in conseguenza altri errori.

La maggior circolazione delle informazioni e i tempi più rapidi consentiti da internet e dall'automazione rendono ancor più critico il problema. Anche se i soggetti interessati dispongono di sistemi informatici privi di dati ridondanti e quindi il numero di correzioni da effettuare sulle basi di dati sia limitato, può essere alto il numero di elaborazioni da ripetere e di procedure, anche amministrative, da mettere in atto per effetto di un unico errore. Occorre quindi approntare procedure che, nel momento in cui si accerti un errore, facilitino la diffusione di "allarmi" alle parti interessate, interne ed esterne, ciascuna delle quali a sua volta dovrà avvertire i propri corrispondenti.

6. Problematiche connesse alla distribuzione

In generale, il compito di rendere disponibili, nei modi e con i requisiti opportuni, i dati pubblici è affidato a **distributori pubblici**, essendo le finalità connesse al diritto della collettività di accedere al patrimonio di conoscenze che i soggetti pubblici producono. Le modalità attuative sono diverse e occorre ragionevolmente regolarle.

1. i distributori pubblici potrebbero rendere conoscibili i dati o direttamente o, tramite contratti di licenza di (ri)diffusione, attraverso **distributori privati**, possibilmente senza esclusiva. La tariffazione per la fornitura dei dati effettuata direttamente dai distributori pubblici dovrebbe tener conto dei soli costi tecnici necessari a renderli conoscibili. La conoscibilità dei dati essenziali dovrebbe rimanere possibilmente gratuita, almeno per qualche forma;
2. la licenza per la (ri)diffusione di dati pubblici da parte di soggetti privati dovrebbe essere affidata sulla base di gara pubblica. In alcuni casi il contratto potrebbe essere attivo per l'amministrazione. Potrebbero essere posti vincoli riguardo alle tariffe. Per particolari tipologie

di dati, il soggetto pubblico detentore potrebbe essere tenuto ad assicurare (anche) la conoscibilità diretta ad altri soggetti pubblici e agli utenti privati, cittadini e imprese;

3. i distributori privati, attraverso più accordi e l'utilizzo di dati anche non pubblici, hanno la possibilità di migliorare e dare valore aggiunto all'informazione. La tariffazione delle licenze di (ri)diffusione, se effettuata da parte dei distributori pubblici, dovrebbe essere tale da non provocare distorsioni della concorrenza e favorire lo sviluppo del mercato dell'informazione;
4. l'eventuale doppia possibilità per la conoscibilità dei dati pubblici, attraverso soggetti sia pubblici sia privati, potrebbe essere fonte di conflitti che occorre evitare; si dovrà anche prevenire che insorgano difficoltà connesse con la validità legale dei dati. I soggetti pubblici devono perciò conservare il diritto patrimoniale sui dati pubblici senza, tuttavia, essere responsabili della qualità nel caso di errori commessi dai distributori privati;
5. per la soluzione dei possibili conflitti tra operatori, sia pubblici sia privati, e come controllore della qualità dei dati potrebbe essere creato uno specifico organo di concertazione e autoregolazione, al quale dovrebbero partecipare i distributori sia pubblici sia privati.

7. Raccomandazioni

La formulazione di una politica per la conoscibilità dei dati pubblici, in un contesto di fruibilità telematica di dati digitalizzati, risulta improcrastinabile.

Sulla base delle considerazioni fin qui svolte, sono esposte di seguito alcune linee per possibili iniziative da assumere da parte del legislatore. Entro la cornice così definita, ogni soggetto pubblico potrebbe poi precisare le specifiche norme per la conoscibilità dei dati pubblici in suo possesso. Le raccomandazioni sono raggruppate per argomento e rappresentano uno stimolo per ulteriori approfondimenti. Le proposte sono riferite esclusivamente al contesto della conoscibilità telematica dei dati pubblici.

Classificazione dei dati pubblici

1. individuazione dei dati pubblici da considerare essenziali;
2. emanazione di indirizzi normativi per assicurare la conoscibilità dei dati essenziali;
3. emanazione di regole tecniche che impediscano, o rendano difficile, la costruzione di archivi derivati che possano violare l'anonimato; in particolare, le regole tecniche devono contribuire ad innalzare il livello dei mezzi ragionevoli;
4. emanazione di regole per l'accesso ai dati sensibili con criteri rigorosi;

Conoscibilità dei dati pubblici

1. indicazione, da parte dei soggetti pubblici che li detengono, dei dati pubblici che sono resi conoscibili e delle relative modalità (accesso, comunicazione, diffusione), in relazione alla natura dei dati stessi e dei soggetti fruitori;
2. individuazione di modalità di conoscibilità dei dati che non penalizzino alcune categorie di utenti rispetto ad altre;
3. individuazione di modelli per la conoscibilità dei dati essenziali, fra i quali:
 - possibilità di consultazione, verifica e, se necessario, aggiornamento dei dati di carattere personale da parte dei titolari;

- forme di facilitazione della ricerca e dell'accesso attraverso le tecnologie web, per la conoscibilità dei dati anonimi;
- 4. definizione di politiche di accesso, di regole di organizzazione dei siti e di livelli di servizio da assicurare ai fruitori;
- 5. eventuale introduzione di servizi di *call center*, per assicurare ai fruitori assistenza tecnica tempestiva e di elevata qualità;
- 6. adeguata pubblicità delle condizioni per conoscere dati non essenziali (responsabilità, orari, livelli di servizio, assistenza, sicurezza, eccetera) e delle loro caratteristiche di qualità;

Soggetti distributori di dati pubblici

1. indicazione delle tipologie di dati pubblici resi conoscibili e fra essi di quelli individuati come essenziali e dei soggetti pubblici che li distribuiscono;
2. indicazione dei criteri di tariffazione per la conoscibilità dei dati;
3. indicazione dei principi in base ai quali definire le licenze di (ri) diffusione dei dati pubblici;
4. effettuazione di gare pubbliche per la (ri) diffusione di dati pubblici da parte di soggetti privati;
5. validità legale delle rappresentazioni digitali dei dati rispetto a quelle tradizionali, da parte di distributori sia pubblici sia privati;

Sicurezza dei dati

1. Esigenza per i distributori pubblici e privati di:
 - adottare misure di sicurezza interne per assicurarne la disponibilità e garantire la riservatezza e l'integrità dei dati pubblici;
 - formulare le misure di sicurezza richieste agli utenti esterni in relazione alle modalità di conoscibilità dei dati;
 - nominare un responsabile della sicurezza dei dati pubblici;

Qualità dei dati

1. emanazione di norme in materia di qualità dei dati pubblici;
2. divieto di acquisizione di dati che siano già in possesso del soggetto pubblico (possibilmente nel suo profilo generale), attraverso l'inversione del ciclo di trasmissione;
3. indicazione di termini temporali per l'aggiornamento dei dati pubblici dichiarati conoscibili rispetto ai tempi di raccolta degli stessi.

Appendice 1: Glossario ²

Accessibilità - Proprietà dei sistemi informatici di essere fruibili senza discriminazioni.

Accesso (giuridico) ai dati - La normativa vigente prevede l'accesso agli atti amministrativi (legge n 241/1990, art. 25, comma1) e l'accesso ai dati personali (legge n. 675/1996, art. 13).

Accesso (informatico) ai dati - Possibilità di utilizzo che permette direttamente al soggetto di esprimere le esigenze e di fruire dei dati. Va dalla sola visualizzazione alla possibilità di formulare interrogazioni o ricerche, all'utilizzo (elaborazione, trasferimento), fino all'aggiornamento (inserimento, modifica e/o cancellazione). L'utilizzo è solitamente predeterminato ed eventualmente associato a profili di *autorizzazione* di singoli utenti o intere classi di utenti. L'iniziativa è per lo più di chi accede e gli accessi possono essere effettuati sia da persone sia da sistemi.

Accuratezza sintattica - Vicinanza del valore del dato ad un valore che viene considerato corretto.

Accuratezza del significato - Vicinanza del valore del dato al valore che assume nella realtà.

Autenticazione - Processo attraverso cui, in una comunicazione tra due parti (uomo-elaboratore, elaboratore-elaboratore) una parte verifica la veridicità dell'identità asserita dall'altra. L'autenticazione si basa in genere sulla verifica del possesso di alcune proprietà che si conosce essere possedute dalla controparte.

Apprendibilità - Impegno richiesto agli utenti per comprendere le possibilità d'impiego delle tecnologie al fine della fruizione delle informazioni desiderate, attraverso le modalità di accesso, comunicazione o diffusione.

Autorizzazione - prevede l'assegnazione di un *profilo di accesso* all'utente (per esempio autorizzazione alla sola lettura, alla lettura e all'aggiornamento, alla cancellazione, alla stampa, eccetera) e il controllo della congruità di ciascuna richiesta rispetto alle tipologie ammesse, ad ogni accesso.

Call center - Punto di accesso ad un insieme di funzioni di assistenza, di norma telefonico.

Classe di dati - Insieme di dati caratterizzati da proprietà comuni, denotabile con un nome.

Completezza - Estensione con cui un insieme osservabile è rappresentato in un insieme di dati. Ad esempio, per un insieme di impiegati di un'organizzazione, un'anagrafica è completa quando sono rappresentati tutti gli impiegati.

Comprensibilità - Impegno richiesto agli utenti per capire il funzionamento degli strumenti, basati su tecnologie informatiche, al fine di utilizzarli per l'accesso o la comunicazione di dati.

Comunicazione - Trasmissione di dati ad uno o più destinatari predeterminati; può essere unidirezionale o bidirezionale; può essere preceduta da una *richiesta* oppure no. L'iniziativa può essere del soggetto che detiene i dati oppure dell'interessato che ne ha fatto richiesta. Il processo di comunicazione di dati digitali può essere totalmente o parzialmente telematico.

Conoscibilità (di un insieme di dati) — Insieme delle regole che determinano la possibilità di fruire dei dati da parte di soggetti interessati.

Consistenza interna (di un insieme di dati) - Rispetto dei vincoli di significato che legano logicamente l'insieme dei dati. Ad esempio, la coppia di dati <CAP, 35122>, <CITTÁ, Roma>, che

² Sono riportati in corsivo i termini per i quali è fornita una definizione nell'ambito del glossario

rappresenta codice di avviamento postale e città di un indirizzo toponomastico, non è caratterizzata da consistenza interna perché non è rispettata la corrispondenza tra codice e città.

Consistenza esterna (dello stesso dato in diversi insiemi di dati) - Proprietà per cui il dato è rappresentato con lo stesso elemento nei diversi insiemi. La consistenza esterna, così come definita, assume che il vincolo di significato cui siamo interessati sia l'identità. La consistenza esterna può essere persa nei processi di trasformazione che il dato subisce, dando luogo alla produzione di copie dello stesso dato non coincidenti.

Contenuto (informatico) - Insieme dei dati che possono essere resi disponibili da un programma applicativo, nel formato in cui sono rappresentati.

Dati esposti - Dati pubblici resi conoscibili a fruitori che possono effettuare accessi agli stessi.

Dato - Rappresentazione di un fenomeno della realtà di interesse in un formato codificato, in modo tale da essere memorizzabile ed elaborabile mediante sistemi informatici.

Dato anonimo - Dato che, in origine o a seguito di elaborazione, non può essere associato a un soggetto identificato o identificabile.

Dato arricchito - Dato risultante da elaborazioni anche basate sul confronto con altre informazioni di differente provenienza, relative allo stesso soggetto. Sono di questa natura, ad esempio, i dati statistici.

Dato di base - Dato sottoposto al trattamento necessario per renderlo elaborabile al di fuori di un singolo sistema o tecnologia.

Dato elementare - Dato che rappresenta un aspetto della realtà di interesse non ulteriormente scomponibile.

Dato grezzo - Dato che si trova nella forma in cui è stato acquisito inizialmente.

Dato essenziale - Dato pubblico di cui cittadini, imprese o altri operatori privati devono poter disporre per esercitare i propri diritti. Non esiste ancora una definizione giuridicamente rilevante di dato essenziale.

Dato personale - Dato relativo a persona fisica, persona giuridica, ente o associazione, identificati o identificabili, anche indirettamente mediante riferimento ad altre informazioni.

Dato proveniente da pubblici registri, elenchi, atti o documenti conoscibili da chiunque - Dato che appartiene ad una collezione tenuta o formata da uno o più soggetti pubblici, in virtù di una norma di legge o di regolamento e che sia conoscibile da chiunque senza condizioni.

Dato pubblico - Dato detenuto da soggetti pubblici perché raccolto o utilizzato da soggetti pubblici, nell'ambito dei propri fini istituzionali.

Dato sensibile - Dato personale idoneo a rivelare l'origine razziale ed etnica, le convinzioni religiose, filosofiche o di altro genere, le opinioni politiche, l'adesione a partiti, sindacati, associazioni o organizzazioni a carattere religioso, filosofico, politico o sindacale, nonché a rivelare lo stato di salute e la vita sessuale.

Dato statistico - Dato elaborato attraverso funzioni di aggregazione su dati elementari o statistici.

Diffusione - Divulgazione ad una platea indeterminata di soggetti anche tramite pubblicazione, tradizionale o su internet. L'iniziativa è di chi detiene i dati. Il grado di effettiva disponibilità dipende dalle modalità prescelte.

Disponibilità (nel contesto della sicurezza informatica) - Capacità di un sistema informatico di assicurare il proprio funzionamento contrastando efficacemente ogni minaccia sia di tipo accidentale sia di tipo intenzionale. Si misura in tempo di regolare funzionamento, calcolato in percentuale rispetto a una finestra temporale predefinita.

Distributore privato - Soggetto giuridico di natura privata che rende disponibili, usualmente a titolo oneroso, dati pubblici ovvero dati arricchiti a partire da dati pubblici.

Distributore pubblico - Soggetto giuridico di natura pubblica che rende disponibili a titolo gratuito o oneroso, dati pubblici.

Dominio di un'amministrazione (nel contesto delle tecnologie informatiche e telematiche) - Insieme delle risorse hardware, software e dei dati che appartengono all'amministrazione.

Dominio di definizione (di un insieme di dati) - Insieme dei possibili elementi cui appartiene un dato.

Elemento - Codifica o formato codificato che può essere assunto da un dato.

Esattezza - Qualità per cui il dato rappresenta correttamente il fenomeno.

Formato codificato - Rappresentazione (di un dato) in forma elaborabile mediante sistemi informatici.

Identificazione - Processo attraverso il quale una risorsa dichiara la propria identità nell'ambito di un sistema o di un'applicazione.

Integrità - Inalterabilità diretta o indiretta del dato, sia da parte di soggetti non autorizzati sia a seguito di eventi accidentali.

Interfaccia utente - Insieme di simboli, comandi, convenzioni attraverso cui l'utente e il sistema informatico interagiscono nella formulazione delle esigenze e nella comunicazione del dato.

Metadato - Proprietà di un insieme di dati. In generale, i metadati sono asserzioni su altri dati che definiscono delle relazioni. La distinzione tra dati e metadati non è intrinseca ma dipende dall'utilizzo. Ad esempio, il nome di un autore può essere un metadato nel contesto della descrizione di risorse elettroniche ed un dato in un sistema di informazioni bibliografiche.

Modalità di scambio (dei dati) - Tecnica attraverso la quale si rende fruibile un dato conoscibile. Le modalità possibili sono: accesso, comunicazione, diffusione.

Nome (del dominio di definizione dei dati) - Denotazione di un dominio mediante una o più parole in linguaggio naturale.

Non ripudio (nell'invio o ricezione di un dato) - Proprietà per la quale il mittente/il ricevente di un dato non possono, in una fase successiva, negare di aver inviato/ricevuto il dato.

Operabilità - Facilità di utilizzo degli strumenti impegno richiesto agli utenti per accedere ai dati attraverso le tecnologie.

Pertinenza e non eccedenza - Qualità rapportate al fine per cui viene posto in essere un trattamento e collegate al rapporto tra dati e adempimenti

Porta applicativa - vedi **Sistema di interfaccia**

Programma applicativo - Insieme delle funzionalità messe a disposizione dell'utente per soddisfare le sue esigenze.

Qualità (di un dato) - Caratteristica del dato desiderata dagli utenti.

Registro pubblico - Registro, archivio, albo formato, utilizzato, conservato da una amministrazione pubblica, previsto da leggi o regolamenti, che raccoglie dati connessi all'espletamento delle attribuzioni e dei servizi svolti dall'amministrazione.

Responsabile della qualità dei dati (o data steward) - Soggetto che, all'interno della amministrazione, ha il compito di definire le regole organizzative e di attuare le iniziative volte al continuo miglioramento delle diverse qualità dei dati trattati nella amministrazione.

Responsabile del trattamento - Persona fisica, persona giuridica, pubblica amministrazione altro ente, associazione od organismo preposti dal titolare al trattamento di dati personali

Richiesta - Domanda volta ad ottenere il rilascio di un documento o dato. Essa può essere effettuata in linguaggio naturale oppure tramite l'utilizzo di moduli (cartacei o elettronici). Nel caso di utilizzazione di moduli cartacei, è usualmente prevista una firma autografa. Nel mondo internet può anche consistere nella sottomissione del proprio indirizzo di posta elettronica al fine di ricevere un dato sotto forma di *file*, *newsletter*, eccetera; oppure del proprio recapito o di fax per ricevere dati su supporto cartaceo.

Riservatezza - Proprietà per la quale nessun utente deve poter ottenere o dedurre dal sistema un dato che non ha il diritto di conoscere.

Schema di dati - Descrizione di un insieme di classi di dati e delle relazioni che intercorrono tra di essi.

Segreto d'ufficio - Vincolo che impegna pubblici ufficiali, dipendenti pubblici e incaricati di servizi pubblici di mantenere il segreto su fatti conosciuti in ragione del loro ufficio (Codice di procedura penale, art. 201);

Segreto di Stato - Vincolo che concerne gli atti, i documenti, le notizie, le attività e ogni altra cosa la cui diffusione sia idonea a recar danno alla integrità dello Stato democratico (L. 801/1977, art. 12).

Segreto professionale - Riguarda una categoria molto ampia di soggetti (ministri di confessioni religiose, avvocati, consulenti tecnici, notai, medici, chirurghi e tutti coloro che svolgono una professione sanitaria); vale anche i giornalisti relativamente ai nomi delle persone da cui hanno avuto notizie in via riservata (Codice di procedura penale, art. 200).

Segreto statistico - Insieme di disposizioni per le quali i dati raccolti da parte degli uffici di statistica nell'ambito di rilevazioni comprese nel programma statistico nazionale non possono essere esternati se non in forma aggregata, in modo che non se ne possa trarre alcun riferimento individuale. I dati così raccolti possono essere utilizzati soltanto per scopi statistici e non possono essere comunicati, se non in forma aggregata o secondo modalità che li mantengano anonimi in rapporto a mezzi considerati ragionevoli, ad alcun soggetto esterno, pubblico o privato, né ad alcun ufficio della pubblica amministrazione. Non rientrano tra i dati tutelati dal segreto statistico gli estremi identificativi di persone o di beni, o gli atti certificativi di rapporti, presenti nei pubblici registri (decreto legislativo n. 322/1989 e codice deontologico da emanarsi in applicazione delle disposizioni del decreto legislativo 281/1999).

Sicurezza - Insieme delle misure (di carattere organizzativo e tecnologico) tese ad assicurare a ciascun utente autorizzato (e a nessun altro) i servizi previsti per l'utente stesso o la categoria di sua appartenenza, nei tempi e secondo le modalità previste.

Sistema di gestione della cooperazione - Insieme di regole organizzative e servizi di gestione che rendono disponibili all'insieme delle pubbliche amministrazioni dati e servizi necessari alla cooperazione tra sistemi appartenenti a domini diversi.

Sistema di interfaccia (o porta applicativa) - Insieme di componenti applicative mediante le quali un'amministrazione espone i servizi applicativi e i dati che intende mettere a disposizione di altre amministrazioni, rendendoli in questo modo accessibili, in condizioni di sicurezza, agli altri domini.

Sistema distribuito cooperativo - Insieme di componenti applicative facenti parte della rete unitaria e della rete nazionale delle pubbliche amministrazioni che rendono possibile lo scambio di servizi applicativi e di dati tra amministrazioni e l'accesso, da parte di una o più amministrazioni, a servizi applicativi e dati resi disponibili da altra amministrazione.

Soggetto pubblico - Amministrazione rientrante fra quelle di cui all'art. 1, comma 2, del decreto legislativo n. 29/1993.

Tempestività - Caratteristica del dato per cui un mutamento del fenomeno al quale il dato si riferisce porta ad un suo aggiornamento in tempo utile rispetto all'utilizzazione.

Titolare del dato pubblico - Pubblica amministrazione che, in base a una norma di legge o regolamentare, ha la responsabilità della raccolta del dato e del suo trattamento.

Titolare del trattamento (di dati) - Persona fisica, persona giuridica, pubblica amministrazione o altro ente, associazione od organismo ai quali competono le decisioni in ordine alle finalità ed alle modalità del trattamento di dati personali, ivi compreso il profilo della sicurezza.

Usabilità (di un dato da parte di un soggetto) - Facilità con cui un soggetto interessato a conoscere il dato riesce ad accedervi, con le tecnologie disponibili e a partire dalla sua situazione fisica, psichica e culturale.

Utente "normale" - Utente di un sistema informatico caratterizzato da un grado di cultura complessiva e da una conoscenza di base degli strumenti informatici tali da far ritenere che lo sforzo necessario per aumentare la *usabilità*, sul versante sia delle tecnologie sia della formazione e addestramento, risulti limitato.

Utente con divario culturale - Utente di un sistema informatico caratterizzato da limitazioni nella cultura, nel linguaggio, nella capacità di utilizzo delle tecnologie dell'informazione che impedirebbero l'accesso a molte tipologie di informazioni rese fruibili a un utente "*normale*". L'utilizzo di tecnologie ad alto grado di usabilità può permettere l'effettivo accesso.

Utente disabile - Utente di un sistema informatico caratterizzato da limitazioni fisiche o psichiche che rendono indispensabile l'utilizzo di specifiche tecnologie assistive per poter accedere al dato.