La qualità dei dati e l'anagrafe digitale

Siamo entrati in una fase di progresso nel raggiungimento della qualità dei dati e della loro integrazione, e può essere interessante ricordare un esempio storico, quasi un'utopia preziosa e illuminante, un approccio visionario, dell'epoca dei primi mainframe monolitici, quando non c'erano personal computer, reti, internet, né smartphone.

Si tratta di un disegno ideato nel 1962 da Bruno de Finetti, ordinario di Calcolo delle probabilità presso l'Università di Roma. Nella figura seguente si ritrovano Enti e gran parte della Pubblica amministrazione di quei tempi con una centralità delle anagrafi dei Comuni e dell'Istat, istituzioni del mondo fiscale, gestori dell'energia e delle comunicazioni, banche, assicurazioni, ecc..

Emerge dal disegno un'attenzione particolare all'anagrafe dei Comuni, percepito come elemento essenziale di unificazione e raccordo dei diversi fenomeni sociali dello Stato (da qui la parola Statistica, citata accanto ai Comuni). Molte delle considerazioni che seguiranno saranno quindi orientate, esplicitamente e implicitamente, al tema anagrafico, punto primario dell'Agenda digitale, col nome di Anagrafe unica, connesso col tema dell'identità digitale. Molte delle considerazioni varranno anche per l'unificazione in genere di archivi, in un'ottica di ottimizzazione, riuso di dati, eliminazione delle ridondanze. Commentava de Finetti, nell'articolo "Il cittadino in codice" di Emanuele Lombardi (marzo 1965): "Secondo questo progetto di anagrafe centralizzata, tutti i dati anagrafici concernenti gli individui, gli enti, ecc. sarebbero registrati nelle memorie di macchine elettroniche e verrebbero *continuamente aggiornate* in base alle *nuove* informazioni e sarebbero quindi *accessibili* istantaneamente per tutti gli usi...". L'articolo prosegue sottolineando più volte il problema dell'aggiornamento in tutti i settori. Ritorneremo su queste parole "aggiornamento", "accessibilità", "nuove informazioni" per collocarle, assieme ad altri termini, nel quadro completo del concetto di qualità dei dati.

Il problema dell'integrazione di basi dati, costruite "verticalmente", a "silos", è un problema generale, che riguarda la maggioranza dei domini applicativi automatizzati: i dati anagrafici e quelli territoriali, i dati bancari, dell'istruzione, degli occupati, dei pensionati o pensionabili, del bilancio pubblico e privato, delle spese, dei debiti e dei crediti, giudiziali, dei contratti, delle imprese, dei dati sanitari come il Fascicolo Sanitario Elettronico, ecc..

Un modello di qualità dei dati per l'era dell'integrazione

Si riporta di seguito il modello di qualità dei dati definito nello standard internazionale ISO/IEC 25012 *"Data quality model"* del 2008, divenuto norma italiana UNI ISO/IEC 25012 nel 2014.

Il raggiungimento di qualità in ogni caratteristica può essere di ausilio per facilitare l'integrazione di basi dati di sistemi diversi che mirano ad integrarsi. L'ISO/IEC 25012 fornisce un elenco di caratteristiche di qualità utili nelle fasi di costruzione di specifiche di progetto, nella definizione di un piano di misurazione, non implica l'adozione di un processo specifico, ma suggerisce soluzioni che possono agevolare il raggiungimento dei livelli di qualità desiderati.

La norma categorizza gli attributi di qualità in quindici caratteristiche considerate da due punti divista: *inerente i dati e dipendente dal sistema.*

Per *inerente* si intende il caso in cui la qualità del dato si riferisce alle sue proprietà intrinseche, a prescindere dal supporto di rappresentazione e dagli aspetti tecnologici.

Per *dipendente dal sistema* si intende il caso in cui la qualità del dato è ricercata all'interno di un sistema informatico che può influenzarla.

Alcune caratteristiche sono interessate da un punto di vista, altre entrambi come di seguito elencate:

Inerente

- accuratezza, intesa come perfetta rispondenza del dato con il mondo che rappresenta;
- attualità, cioè del giusto tempo con il quale il dato è creato o utilizzato;
- coerenza, non contraddittorietà con altri dati, univocamente interpretabili;
- completezza, ove tutti gli attributi necessari sono presenti, con tutte le fonti;
- credibilità, per dati provenienti da fonte certa.

Inerente e dipendente dal sistema

- accessibilità, il dato è accessibile in varie situazioni, anche da disabili;
- comprensibilità, il significato del dato (e del metadato) è chiaro e immediato;
- conformità, il dato risponde prioritariamente a regolamentazioni, anche locali;
- efficienza, il dato è gestito con risorse accettabili e tempi adeguati allo scopo;
- precisione, il dato possiede il livello di misura discriminante necessario;
- riservatezza, il dato può essere utilizzato solo da utenti autorizzati;
- tracciabilità, gli accessi al dato sono registrati.

Dipendente dal sistema

- disponibilità, il dato è disponibile e re-interrogabile;
- portabilità, il dato è gestibile e può migrare da un ambiente all'altro;
- ripristinabilità, il dato è salvato in un ambiente sicuro ed è recuperabile.

Occorre accennare che la qualità e integrazione dei dati dipende da vari fattori, tra cui:

- adesione ad un modello comune di qualità;
- supporto e presenza di software specifico ad alta qualità;
- supporti statistici anche per dati non strutturati (es. *Big data*);
- tecnologie avanzate (es. *Cloud*);
- tecniche di integrazioni virtuali di dati;
- supporto normativo/legislativo;
- presenza di indirizzi strategici e linee guida;
- diffusione di una cultura della condivisione inter-amministrativa di dati;
- aggiornamento delle regolamentazioni procedurali;
- interventi organizzativi e responsabilità identificate;
- coinvolgimento del cittadino (es. Open data individuale e collettivo/statistico, web);
- disponibilità di tecnologie efficienti (es. Larga banda).

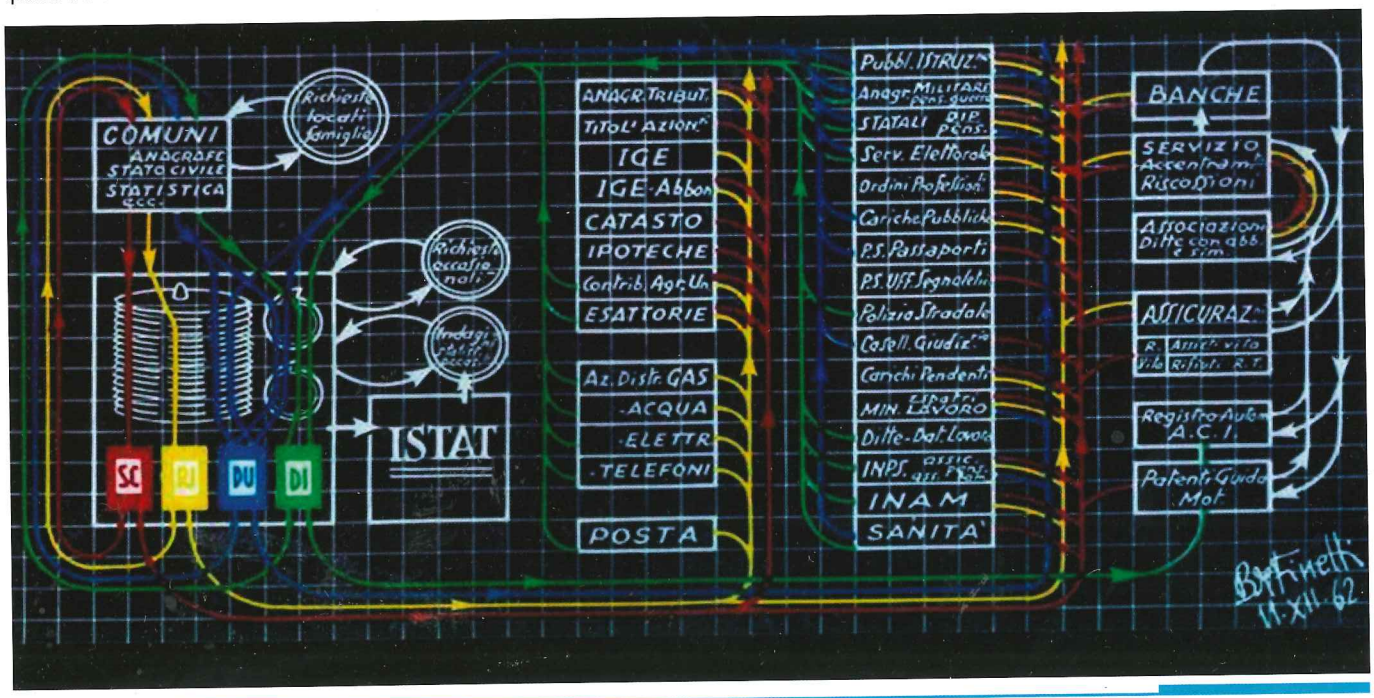

Figura 1 - Disegno ideato da Bruno de Finetti - Per gentile concessione di Fulvia de Finetti da www.brunodefinetti.it

Limitandoci al primo punto (adesione ad un modello comune di qualità), riprendiamo alcune caratteristiche di qualità *inerente* il dato.
L'accuratezza, intesa come perfetta rispondenza del dato al mondo che rappresenta, è probabilmente la caratteristica che meglio sintetizza tutte le caratteristiche puramente inerenti. Il dato rispecchia la realtà, quindi è vero, è l'ultimo esistente, è coerente, è completo, è credibile. Dire che un dato è accurato, secondo la norma, vuol dire, per dirla con sinonimi, che è esatto, corretto, ecc..
Per la verifica di accuratezza occorre una base dati certificata con l'elenco dei dati reali e la capacità di contestualizzare la corrispondenza del dato alla realtà. Occorre la possibilità di accedervi, di tenerla aggiornata, di evitare duplicazioni esterne presto obsolete, e così via.
Circa il problema delle duplicazioni andrebbero eliminate nuove rilevazioni ad hoc di indirizzi (ad esempio da parte di banche, gestori di energia, ecc.), specialmente cartacee o telefoniche, senza verifica in linea e particolari accorgimenti, che non fanno altro che aumentare il rischio di entropia, per non dire il "caos", dei dati. L'archivio anagrafico unificato, ad esempio, non è solo un agglomerato di dati desunti dalle varie fonti (forse con date di aggiornamenti diversi), ma una base di dati validata e aperta ad un canale certificato di modifica o conferma da parte dell'utente, fonte credibile di confronto, accessibile con tecnologie moderne (call/contact center, web, ecc.), supportato da tecnologie *Cloud, Big data, Open data* e che dispone per taluni servizi di larga banda. Il dato individuale è necessariamente *Open*, per l'utente finale che potrà vedere i propri dati presenti nel sistema, ed eventualmente poter attivare opportune azioni.
L'archivio "centrale", di tipo virtuale o fisico, sarà il riferimento riguardante gli aspetti più rilevanti della vita nazionale: liste elettorali, fisco, pensioni, assicurazioni, dati dei comuni, regioni, settore ospedaliero, fonti energetiche, comunicazioni, settore privato, ecc.. Tutti attingono da un unico data base certificato e tutti contribuiscono, con adeguati controlli e titoli credibili di autorizzazione, alla sua alimentazione e correzione. Il presupposto di un archivio unico centralizzato è quello della modifica autorizzata/certificata e della sua mutabilità, con coerenza e completezza delle informazioni, tempestività di aggiornamento, verifiche di accuratezza semantica contestuale. In pratica l'archivio sarà soggetto a rispondere a tutte le caratteristiche di qualità del modello adottato.
Si può comporre il quadro di riferimento che segue in cui si ritrovano quattro elementi essenziali in un'ottica di integrazione: un sistema *informativo/informatico A*, preso ad esempio, che gestisce dati per *utenti intermedi e finali*, il *mondo reale* ove sono presenti numerosi sistemi informativi/informatici B, C...Z (basti pensare nel caso anagrafico ai circa 8.100 Comuni), il *call/contact center* di supporto all'archivio centrale unificato per la gestione di chiarimenti e aggiornamenti da parte utente.
Anche per lo stesso *call/contact center* si potrà presumere il rispetto delle caratteristiche di qualità del modello in esame, in particolare per l'accessibilità (e multicanalità), la comprensibilità, l'efficienza, la disponibilità del servizio, nonché per la credibilità dell'identità dell'utente che vi entra in contatto.
Il dialogo tra il sistema e l'utente specifico avverrà sempre più con approcci Open data, sia di tipo individuale ove l'utente rivede la sua completa situazione inter-amministrativa, come si dice ora "con un semplice click", che a livello di collettivo/statistico per informazioni generali pubbliche.
L'uso del web, dei *social nework* in generale, dell'uso del "mobile" potrà svilupparsi maggiormente nel prossimo futuro anche per questioni inerenti gli aspetti di validazione dei dati da parte dell'utente.
L'attualità, cioè la presenza in archivio di dati aggiornati col "giusto tempo", richiederà probabilmente una serie di miglioramenti di processi e regolamenti. In alcuni casi è sufficiente un aggiornamento dei dati nell'ordine dei mesi, in altri casi dei giorni, se non ore, minuti, secondi. Spesso i dati non attuali derivano, nel campo amministrativo, da processi burocratici obsoleti, non efficienti o non regolamentati. L'ideale sarà poter seguire in tempo reale le variazioni anagrafiche, man mano che si verificano.
Spesso c'è un ritardo tra il verificarsi del fenomeno reale rappresentato dal dato e la sua raccolta o memorizzazione nei sistemi. Anche la cancellazione di dati obsoleti rientra nello stesso fattore qualitativo, talvolta causa di successive situazioni non accurate.
La coerenza tra dati, cioè non contraddittori con altri dati e interpretabili in modo univoco, è un problema di grandi dimensioni che coinvolge anche i processi di condivisione delle informazioni e delle norme tecniche adottate. Il tema riguarda i dati con lo stesso nome, ma di contenuto diverso, oppure con nomi diversi, ma riguardanti lo stesso fenomeno.
Si profila una intensa attività di allineamento e un piano di uniformità dei formati di interscambio.
Si apre il discorso di un dizionario dati che contenga metadati, descrizioni e formati standard di interscambio.
Si apre anche il problema della coerenza semantica rispetto al contesto. Un semplice caso di non coerenza, ad esempio, è quando la data di assunzione di un impiegato è antecedente alla data di nascita.
L'attenzione alla coerenza apre la porta alla interoperabilità dei sistemi.
La completezza è la presenza di tutti gli attributi necessari, ed implicitamente di tutte le fonti utili. Un esempio di non completezza può essere un indirizzo anagrafico senza il codice postale, un numero civico senza la scala o il piano e l'interno, un riferimento personale senza il numero di telefono, l'assenza di una data di scadenza.
La completezza riguarda anche i record di un archivio, come ad esempio l'elenco dei cittadini residenti, dei cittadini immigrati, dei cittadini residenti all'estero.
La credibilità riguarda il dato proveniente da fonte certa ed ufficiale. Essa è alla base di tante considerazioni sulla comunicazione ed il supporto alle decisioni. L'archivio anagrafico unificato deve essere fondamentalmente credibile, rappresentando la fonte primaria del dato vero da distribuire a tutti.
La conformità, tra le altre caratteristiche del modello ISO/IEC 25012 che non possiamo in questa sede tutte approfondire, merita un particolare accenno, in quanto si precisa che eventuali Linee guida o Leggi vigenti in Italia, hanno un valore prioritario rispetto al modello.
Si cita all'uopo principalmente il DPCM del 23 agosto 2013 che istituisce l'Archivio Nazionale della Popolazione Residente (ANPR).
In esso sono citate varie caratteristiche di qualità (dei dati e del software) come ad esempio: la disponibilità dei dati, l'accessibilità, la riservatezza, la tracciabilità, la sicurezza, l'interoperabilità, la coerenza. Si intravvedono inoltre varie relazioni specifiche delle caratteristiche del modello di qualità ISO con alcuni temi regolamentati, molto attuali, come l'accessibilità prevista dalla Legge 4/2004, la riservatezza con il Codice della privacy, come altri citati nelle Linee guida dell'Agid, Agenzia per l'Italia digitale.
Si sottolinea in particolar modo anche che la norma ISO/IEC 25012 è oggetto della determinazione

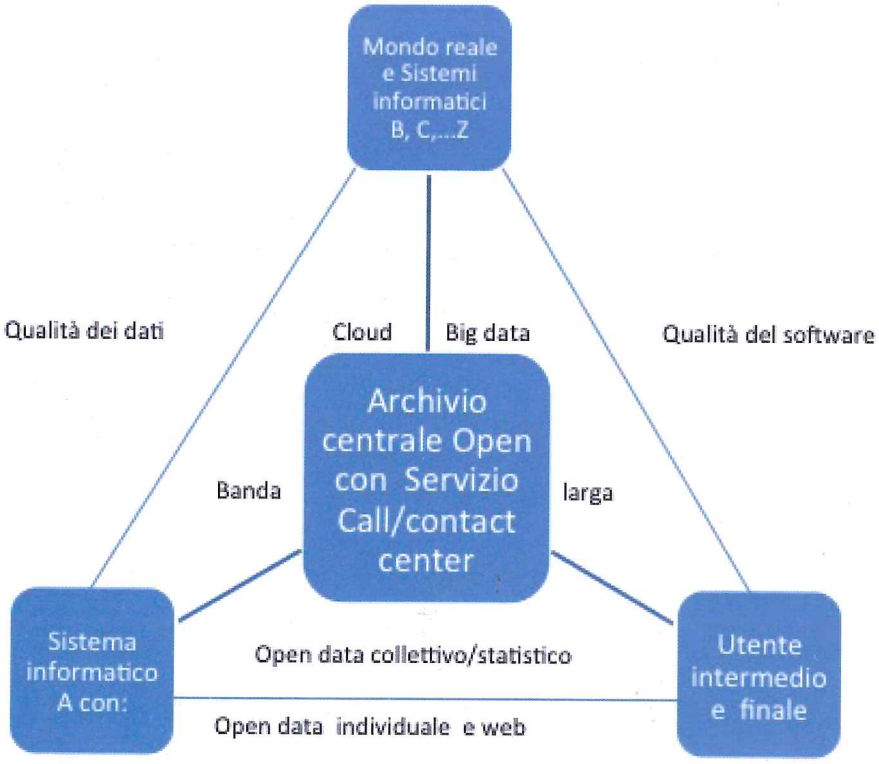

Figura 2 - Elementi di un archivio centrale unificato

commissariale n. 68/2013 dell'Agid che raccomanda l'osservanza di quattro caratteristiche di qualità inerenti (accuratezza, attualità, coerenza, completezza) nella gestione di varie banche dati critiche di interesse nazionale, tra cui la stessa anagrafe nazionale della popolazione residente, già indicate dall'art. 60 del Codice di Amministrazione Digitale.

Conclusione

Aver affrontato le caratteristiche di qualità suddette ci porta a sottolineare una demarcazione non irrilevante dell'uso degli standard in generale: alcuni trattano di caratteristiche rilevanti soggette a leggi o regolamentazioni nazionali, quindi "obbligatorie" in modo esplicito o implicito; in tale caso l'appellativo di norme tecniche è quanto mai appropriato. Altri standard internazionali, o parti di essi, costituiscono utili indicazioni, la cui adesione è "volontaria", da considerarsi comunque importanti per il miglioramento della qualità dei prodotti e dei servizi: anch'esse rappresentano una spinta al progresso della società, essendo espressione delle migliori prassi industriali, accademiche e delle amministrazioni di tutto il mondo.

Circa il tema dell'integrazione si stanno verificando molti segnali positivi che conducono alla interoperabilità dei sistemi ed alla integrazione delle basi dati, a beneficio delle amministrazioni e degli utenti.

Il CAD, Codice dell'Amministrazione Digitale, stabilisce che "le pubbliche amministrazioni comunicano tra loro attraverso la messa a disposizione a titolo gratuito degli accessi alle proprie basi dati...", facilitando le procedure di interscambio, migliorando la condivisione inter-amministrativa e riducendo le duplicazioni. La predisposizione da parte degli Enti centrali competenti di ulteriori regole tecniche saranno di ausilio al programma di valorizzazione del patrimonio informativo pubblico, specificando nuove regole di comunicazione.

Il maggiore sviluppo di applicazioni di Open data metterà a disposizione dati statistici alla comunità e dati individuali ai titolari (in modo riservato).

In tale contesto è auspicabile un ruolo attivo e collaborativo dell'utente finale nella comunicazione agevolata di informazioni che lo riguardano ad Enti centrali. L'utente potrà sviluppare, con opportuni servizi innovativi, la capacità di contribuire alla qualità dei dati, con conferme o correzioni, per le varie caratteristiche di qualità.

Domenico Natale
Presidente della Commissione UNINFO SC7 Ingegneria del software
Editor della norma UNI ISO/IEC 25012

Profili professionali per la sicurezza delle informazioni

Già quando si parla di "sicurezza delle informazioni" succede spesso che i non addetti ai lavori non inquadrino esattamente il contesto, figuriamoci la situazione quando si tratta di valutare l'assunzione di un soggetto che di questo si occupa all'interno di un'azienda o di inserire un profilo dedicato all'interno di un bando di gara. Ma facciamo un passo indietro...a differenza di settori più tradizionali e consolidati *l'Information & Communication Technology* (ICT) sta solo recentemente iniziando a trovare una sua dimensione stabile all'interno delle aziende moderne, dotandosi di una serie di profili *"standard"* e di una propria dignità indipendente e riconosciuta, anche grazie alle numerose iniziative che in questi anni si sono concentrate sul tema, tra cui la più degna di nota nel nostro continente è certamente *l'European e-Competence Framework*[1].

La *cybersecurity* o sicurezza informatica, che in un'accezione non esclusivamente legata ai sistemi ICT ma a tutte le informazioni che vengono trattate indipendentemente dalla natura degli strumenti impiegati viene più correttamente indicata come sicurezza delle informazioni, è tuttavia rimasta un po' in disparte come testimoniato dalla enorme varietà di titoli disomogenei in cui ci si imbatte esplorando un qualsiasi social network lavorativo. Per questo motivo, dopo il sostanziale recepimento *dell'European e-Competence Framework* nella norma UNI 11506, il sottocomitato 27 di UNINFO che proprio di sicurezza delle informazioni tratta essendo coinvolto nei lavori delle norme ISO/IEC 27000[2] e 15408, ha deciso di cercare di porre fine a questa imbarazzante situazione mettendo mano ai mattoni creati dalla sopracitata norma UNI e costruendo con essi un insieme minimo ma consistente dei profili professionali più comuni e più rilevanti nell'ambito. Questa iniziativa giunge con il finire del 2014 finalmente nelle sue fasi conclusive e i suoi contenuti saranno a breve recepiti da una norma nazionale dedicata a questo scopo e completamente armonica con la UNI 11506 e con le leggi dello stato promulgate in materia. Qualche anticipazione prima dell'inchiesta pubblica che precede la pubblicazione di ogni norma? I profili professionali che sono stati presi in considerazione sono principalmente 9 a cui se ne aggiungono alcuni altri che sono richiamati direttamente da leggi o norme (quali ad esempio il data protection officer legato al futuro regolamento in materia di protezione dei dati personali). Questi profili sono stati catalogati seguendo le linee di ragionamento facilmente intuibili nella tabella sottostante e sintetizzano le principali figure professionali di cui un'organizzazione correttamente strutturata ha bisogno al giorno d'oggi per potersi difendere adeguatamente nel cosiddetto *"cyberspace"* ma anche nei contesti più tradizionali (sicurezza fisica dei documenti cartacei et similia). Rispetto a questo schema vale la pena notare che con ogni probabilità le aziende di più piccole dimensioni tenderanno a sovrapporre questi ruoli mentre quelle

		ATTIVITÀ	
		GESTIONE	VERIFICA
CONTESTO	DIREZIONE	Responsabile della sicurezza delle informazioni (CISO)	Analista di processo per la sicurezza delle informazioni
	ORGANIZZAZIONE	Manager della sicurezza delle informazioni	
	PROCESSI	Specialista di processo della sicurezza delle informazioni	
	APPLICAZIONI	Specialista applicativo della sicurezza delle informazioni	Analista tecnico per la sicurezza delle informazioni
	INFRASTRUTTURA	Specialista infrastrutturale della sicurezza delle informazioni	
	INCIDENTI	Specialista nella risposta agli incidenti	Analista forense

U&C
Unificazione&Certificazione
LA RIVISTA DELLA NORMAZIONE TECNICA

DOSSIER: AGENDA DIGITALE E NORMAZIONE TECNICA

Alimenti bilanciati e a basso costo

Effetti delle vibrazioni sugli edifici

Le nuove norme del glow-wire

1
Gennaio 2015
Anno LX

Poste Italiane s.p.a. - Spedizione in Abbonamento Postale - D.L. 353/2003 (conv. in L. 27/02/2004 n° 46) art. 1, comma 1 – CN/BO