



# SOSTEGNO AI COMUNI NELLA RACCOLTA E PRODUZIONE DI OPEN DATA

L'ESPERIENZA DEL CENTRO INNOVAZIONE E TECNOLOGIE DELLA PROVINCIA DI BRESCIA



# PROGETTO OPEN DATA

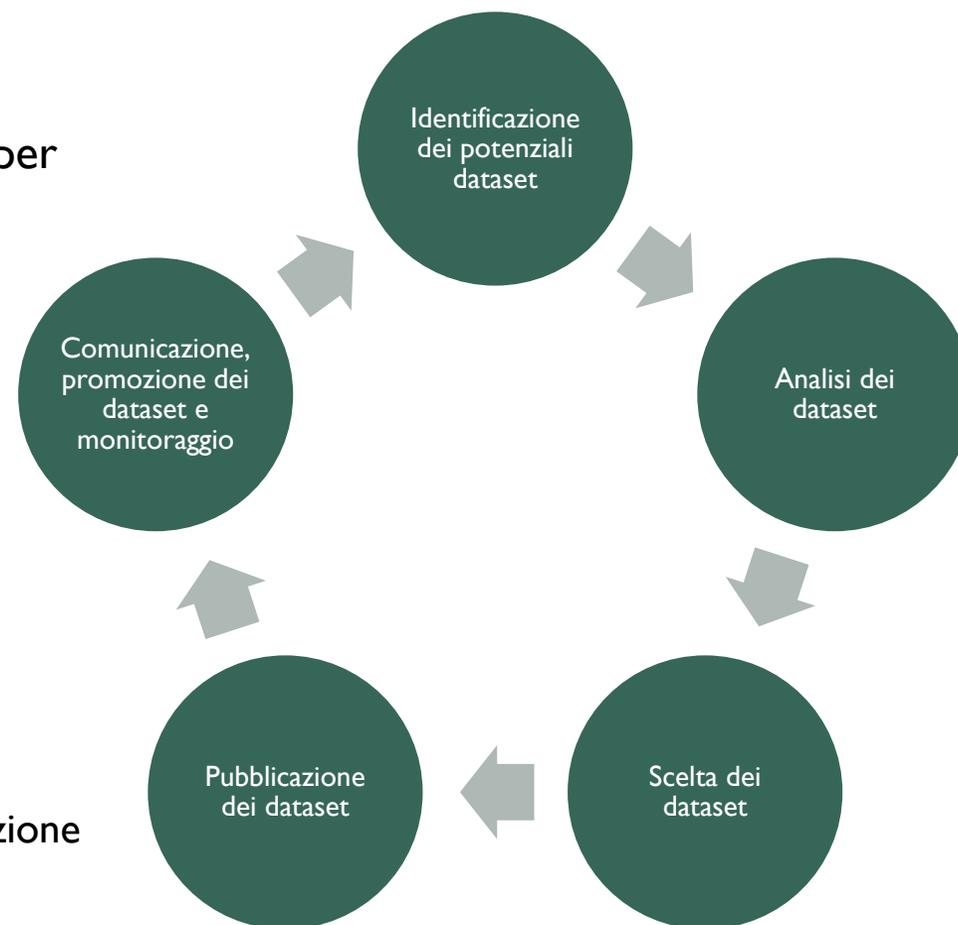
- DGR n. 2585: « approvazione Agenda Digitale Lombarda 2012-2015 »
- Lancio del portale il 23 marzo 2012, all'indirizzo [www.dati.lombardia.it](http://www.dati.lombardia.it)
- Piattaforma basata su SaaS (*Software as a Service*) di Socrata, già riferimento per i *datagov* anglosassoni
- Offre API RESTful, librerie e SDK per i linguaggi più diffusi – al fine di facilitare la realizzazione di applicazioni, agevolando l'interoperabilità e, quindi, l'accesso ai dati per gli utenti
- Utilizza il tool ETL (*Extract, Transform and Load*) Pentaho-Kettle per la pubblicazione e l'aggiornamento
- La Lombardia è fra i primi enti, per numero di dataset, ad alimentare il portale [www.dati.gov.it](http://www.dati.gov.it) – catalogo nazionale dei metadati relativi ai rilasci in formato aperto delle pubbliche amministrazioni



Regione  
Lombardia

# CRITERI REGIONALI DELL'OPEN DATA

- DGR n. 2904: « criteri generali per l'Open Data »
- Si definiscono gli Open Data come dati pubblicati caratterizzati per essere *liberamente utilizzati, riutilizzati e redistribuiti*. Ovverosia:
  - **COMPLETI**: dovranno essere accessibili tutti i dati, fuorché quelli soggetti a valide restrizioni di riservatezza e sicurezza;
  - **PRIMARI**: i dati saranno raccolti alla fonte, con il massimo livello possibile di dettaglio, non in forme aggregate o modificate;
  - **TEMPESTIVI**: saranno resi pubblici tanto velocemente quanto è necessario per preservarne il valore;
  - **ACCESSIBILI**: disponibili al più ampio numero di utenti;
  - **LEGGIBILI DAI COMPUTER**: processabili in modo automatico;
  - **NON DISCRIMINATORI**: disponibili a chiunque, senza registrazione;
  - **NON PROPRIETARI**: disponibili in formato aperto;
  - **LIBERI**: l'uso ed il riuso non dovrà essere soggetto a nessuna restrizione derivante da copyright o brevetto.



# AGENDA DIGITALE LOMBARDA

- « Paniere » di 50 dataset, con relativi standard
- Suddivisi in 10 categorie o aree tematiche (ambiente, commercio, cultura, energia, mobilità e trasporti, sicurezza, statistica, territorio, trasparenza e turismo)
- Selezione sulla base dei dataset ritenuti più utili, frutto sia di un processo collaborativo con gli E.E.L.L. coinvolti sia di monitoraggio degli open data messi a disposizione dalle altre pubbliche amministrazioni italiane
- Non si tratta solo di un elenco di 50 « titoli », ma per ognuno è stato anche proposto un tracciato record di base, con l'obiettivo di:
  - Migliorare il livello di *qualità* dei dataset;
  - Favorire la *completezza* delle informazioni;
  - Permettere il *confronto* tra i dati di enti diversi;
  - Facilitare la *fruibilità* dei dati.



# STANDARD DEI DATASET DEL « PANIERE » REGIONALE

Denominazione campo	Descrizione	Tipo	Obbligo
Anno	Anno di riferimento	Numero	M
Codice_Pratica	Codice identificativo della pratica	Testo	O
Richiedente	Ragione sociale del richiedente	Testo	M
Protocollo	Numero di protocollo	Testo	M
Oggetto_Richiesta	Indicare l'oggetto della richiesta	Testo	O
Data_Richiesta	Data della richiesta	Data	M
Link_pratica	Link alla pratica	Web URL	O

Tabella 6. Tracciato standard Pratiche SUAP

Esempio di standard (dati riferibili alle pratiche SUAP), così come riportato nel « paniere » regionale per gli E.E.L.L.

- **DENOMINAZIONE CAMPO:** intestazione di ciascuna colonna nel formato tabellare di partenza
- **DESCRIZIONE:** chiarificazione, vedi sopra
- **TIPO:** i formati possibili per i dati (da impostare nell'estratto tabellare di partenza) sono i seguenti:
  - Numero;
  - Testo;
  - Web URL;
  - Operatore booleano (spunta);
  - Data;
  - Valuta (non consigliato);
  - Percentuale;
  - Coordinate geografiche (coppia).
- **OBBLIGO:** contrassegnato con le lettere **M** (*mandatorio*) e **O** (*opzionale*) – nel primo caso alcune o tutte le celle della colonna possono rimanere vuote.



# QUALI FORMATI SCEGLIERE PER I DATI?



- La piattaforma [www.dati.lombardia.it](http://www.dati.lombardia.it) implementa una serie di funzionalità avanzate da applicare ai dataset in formato tabellare – quali, ad esempio, l'ordinamento dei dati o i link a pagine esterne
- Perché sia possibile ordinare le celle secondo criterio crescente/decescente, queste devono ospitare dati in formato compatibile con la sopracitata funzione: numeri e dati compatibili con lo standard ISO 8601
- Nessun separatore delle migliaia e punto come segno di separazione dei numeri decimali
- La visualizzazione finale può essere modificata, in fase di procedura di caricamento, dall'interfaccia utente
- Per la rappresentazione geografica il sistema di coordinate (geodetico) è il WGS84 – i dati inseriti potranno essere utilizzati per creare una vista derivativa di tipo « mappa »
- Può essere caricato uno *shapefile* ESRI (per sistemi informativi geografici) come allegato al dataset

# INIZIATIVE DEL C.I.T. DELLA PROVINCIA DI BRESCIA

- DGR n. 7256: « iniziative regionali per la digitalizzazione degli Enti Locali »
- 10 E.E.L.L. aderenti al progetto Open Data a cui il C.I.T. ha offerto assistenza nella produzione e nel caricamento
- 10 dataset per ciascun comune, selezionati fra quelli del « paniere » regionale
  
- La *selezione* ha seguito i seguenti criteri:
  - **GESTIONE DEL DATO**: se detenuto tramite software proprietario (gestionale), sotto forma di database digitale (es: MS Excel, MS Access...), archivio cartaceo o non registrato in nessuna forma;
  - **FREQUENZA DI AGGIORNAMENTO**: tempestiva, mensile, trimestrale o semestrale;
  - **PRIORITÀ DELL'ENTE**: tre livelli di priorità;
  - **POSSESSORE DEL DATO**: il solo comune o anche la Provincia di Brescia;
  - **OBBLIGHI NORMATIVI COLLEGATI**: connessi alla c.d. Amministrazione Trasparente;
  - **NUMERO CAMPI TRACCIATO**: numero minimo di informazioni che devono essere rese disponibili nell'atto di caricamento del dataset sul portale Open Data di Regione Lombardia.



# OPERAZIONI PRELIMINARI



- Registrazione user (Socrata ID) per ciascun E.E.L.L.
- Inserimento del logo comunale (in formato 220x220)
- Creazione di *Application Token* – codice alfanumerico, utilizzato dalle Open Data API in luogo dell'indirizzo IP (o per autenticazione su protocollo di rete OAuth, senza necessità di fornire alcuna password)
  
- A questo punto l'account è pronto per le operazioni di caricamento dei nuovi dataset
- Il nome della collezione di dati in fase di pubblicazione deve rispecchiare lo standard:

Ad esempio: #20 Pratiche SUAP

# CARICAMENTO NUOVA COLLEZIONE DI DATI

## Crea un Nuovo Dataset

Per iniziare, scegli un nome per il tuo dataset

Nome Dataset #20 Pratiche SUAP

Crea Dataset



### Aggiungi dati

Aggiungi tabellare, geospaziale, non analizzabile o dati con link esterno

Aggiungi Dati ▾



### Rivedi e configura i dati

Configura e conferma i tuoi dati prima di pubblicarli

Revisiona i dati



### Edita meta dati del set di dati

Aggiungi contesto per aiutare gli utenti a capire i tuoi dati

Edita i metadati



### Edita metadati della colonna

Descrivi le tue colonne per aiutare gli utenti a interpretare i tuoi dati

Edita i metadati

## + AGGIUNGI DATI



### Trascina Qui un Nuovo File di Dati

...oppure clicca Sfoglia per selezionare un file dal tuo computer

Scorri

I tipi di file supportati sono .csv, .tsv, .xls, .xlsx, .zip (shapefile), .json (GeoJSON), .geojson, .kml, .kmz

I file non analizzabili possono essere caricati anche per l'archiviazione

Il caricamento delle cartelle non è supportato

- *Drag & drop* della collezione di dati per avviare il caricamento sulla piattaforma
- Ampia scelta per tipologia di estensioni supportate – sia formati aperti che proprietari
- Il tool ETL (*Extract, Transform and Load*) provvede all'estrazione dei dati, al loro processo di trasformazione (normalizzazione, traduzione, consolidamento) ed, infine, a memorizzarli nella tabella del sistema di sintesi (la piattaforma Open Data)
- I dati possono essere revisionati per modificarne la *formattazione*, prima del salvataggio

## EDITA METADATI

- I metadati (*data about data*) sono un complemento indispensabile alle collezioni di dati, li censiscono attraverso una serie di parametri standardizzati (per interoperabilità)
- Titolo del dataset (nome della collezione di dati) e breve descrizione del contenuto
- Categoria (vedasi « paniere » regionale degli standard) e tag/parole chiave – strumento che si ricollega al concetto di SEO (*Search Engine Optimization*): visibilità sui motori di ricerca
- Licenza: *Italian Open Data License 2.0*
- E-mail di contatto, affinché sia individuato univocamente un referente
- Frequenza di aggiornamento: tempestiva, mensile, trimestrale o semestrale
- Data ultima modifica dei dati



# ITALIAN OPEN DATA LICENSE 2.0



- « Italian Open Data License » (IODL) è un contratto di licenza il cui scopo è di consentire agli utenti la condivisione, modifica, uso e *riuso* di dati e informazioni contenute nei database – ferma restando la libertà da parte di terzi di fare altrettanto
- Mira a facilitare il riutilizzo delle informazioni pubbliche nel contesto dello sviluppo della società dell'informazione
- I diritti concessi dalla presente licenza non implicano alcun trasferimento di diritto di titolarità sulle banche di dati, sui dati e sulle informazioni pubbliche
- A condizione di includere la fonte delle informazioni e il nome del « licenziante » – ovvero il soggetto (ente, persona fisica o giuridica) che fornisce i dati secondo i termini della presente licenza



# PROCESSO DI MANUTENZIONE DEGLI OPEN DATA

AGGIORNAMENTO AUTOMATICO DEI DATI SUL SISTEMA DI REGIONE LOMBARDIA



# UTILITÀ PER L'AGGIORNAMENTO DEI DATASET

- La procedura di pubblicazione dei dataset nel portale Open Data prevede un primo step di caricamento manuale del file all'interno del profilo degli E.E.L.L. – tuttavia ciò non assolve tutti gli obblighi previsti dal bando di R.L.
- Il loro aggiornamento, costante nel tempo, è considerato un *requisito obbligatorio*
- A tal proposito è stato predisposto da Socrata lo strumento gratuito DataSync 
- Il software (Java) consente la sostituzione automatizzata di una collezione di dati (identificata univocamente da un ID con formato xxxx-xxxx) con una sua versione aggiornata, mantenuta in locale
- L'identità del pubblicatore, ai fini della comunicazione fra software e piattaforma, è verificata tramite App Token

# UTILITÀ PER LA PIANIFICAZIONE

- DataSync può condensare la propria attività in una riga di comando (*command-line command*)
- Quest'ultima è compatibile con diversi strumenti di pianificazione quali, ad esempio, l'Utilità di Pianificazione (*Task Scheduler*) Microsoft – già installata fra le utility di qualsiasi sistema operativo Windows



- In questo modo l'attività può essere programmata in maniera ricorsiva, così da rispettare le scadenze previste da R.L. per quello specifico standard: aggiornamento *mensile*, *trimestrale* o *semestrale*

## SUCCESSIVI STEP DI LAVORO

- Predisposizione dei file .csv per tutti i dataset pubblicati dall'E.L.
- Impostazione di un meccanismo automatico di pubblicazione, con l'ausilio di DataSync, per ciascun dataset
- Impostazione di un'attività ricorsiva all'interno dell'Utilità di Pianificazione di Windows
- Aggiornamenti periodico dei file .csv mantenuti in locale, secondo la scadenza definita dallo standard di R.L.

