

# OPEN DATA NELLA PA

## LA PREPARAZIONE DEI DATASET E DELLE RISORSE

a cura di **Andrea Borruso**

# Open Data nella PA

Preparazione dataset e risorse

# Andrea Borruso

contatti

✉ [andrea.borruso@ondata.it](mailto:andrea.borruso@ondata.it)

🐦 @aborruso

# Glossario

# Glossario

- **Dataset:** è un insieme di dati, un raccoglitore. Spesso si usa però anche per fare riferimento a una intera banca dati o a una tabella
- **Risorsa** (o distribuzione): è l'elemento che "contiene" i dati. Può essere un file, un URL (un "puntamento" a un file o a un servizio)

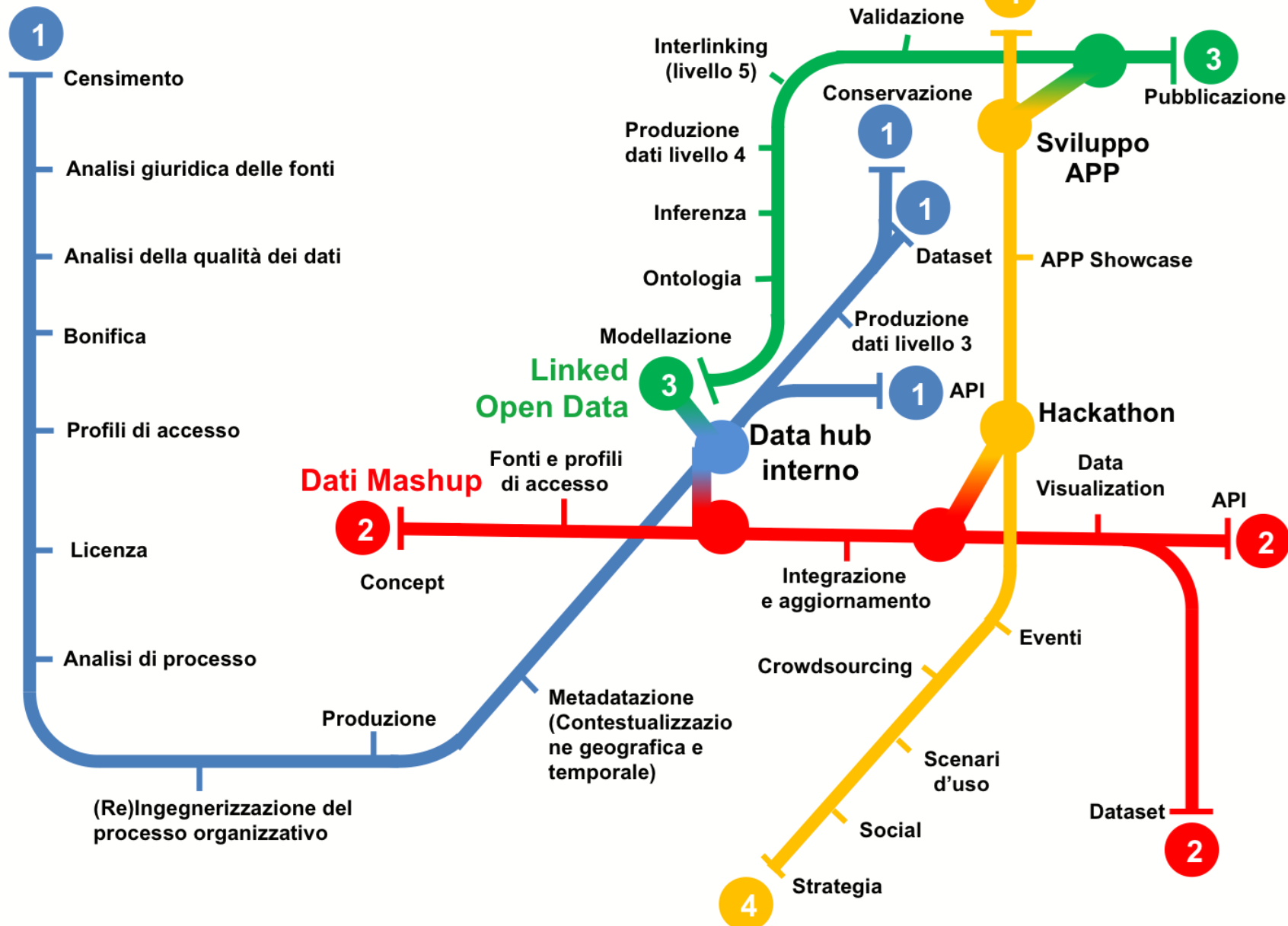
## Glossari utili

1. [Open Data Handbook](#)
2. Le (nuove) [Linee guida open data](#) (e a proposito ...)

# Organizzazione



# Dati Nativi



[Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico](#)

# Organizzazione

*"È necessario definire una chiara data governance interna con l'individuazione di ruoli e relative responsabilità; integrare le sue fasi sia verticalmente, rispetto ai processi interni già consolidati, che orizzontalmente rispetto alle necessità delle diverse amministrazioni."*



# Organizzazione

Fare open data non è un mero adempimento tecnologico:

- l'ottimizzazione dei processi esistenti all'interno dei quali l'Open Data deve far parte integrante;
- la dislocazione di soluzioni interoperabili che possano contribuire all'ottimizzazione dei processi;
- una riduzione nei costi e nei tempi di accesso al capitale informativo

# Organizzazione

Fare open data non deve essere un mero adempimento tecnologico:

- una riduzione della complessità dei processi interni attraverso il consolidamento delle attività derivate dai punti precedenti;
- l'ottimizzazione dei tempi e dei canali di comunicazione istituzionali relativi al capitale informativo verso risorse esterne all'amministrazione.

# Organizzazione

*"Fare in modo che il processo di apertura dei dati **non sia** sempre e solo **parallelo** a quello di gestione dei dati ma **pienamente integrato**."*

*"Individuare una chiara data governance interna con **professionalità strategiche e specifiche**"*

 *once only*

# Organizzazione

## Gruppo lavoro open data

Il gruppo che promuove l'uso e la diffusione degli Open Data. Esso **riporta** all'interno dell'amministrazione le **novità** inerenti il mondo dell'Open Government, media e valuta le **esigenze** di **pubblicazione** dati in base alle **normative** di riferimento, e ne cura la razionalizzazione rispetto agli altri processi di apertura del dato. Ha la responsabilità di **pianificare** e **coordinare** l'evoluzione continua dell'apertura dei dati nell'amministrazione, nonché dell'infrastruttura IT a supporto.

# Organizzazione

## Gruppo lavoro open data

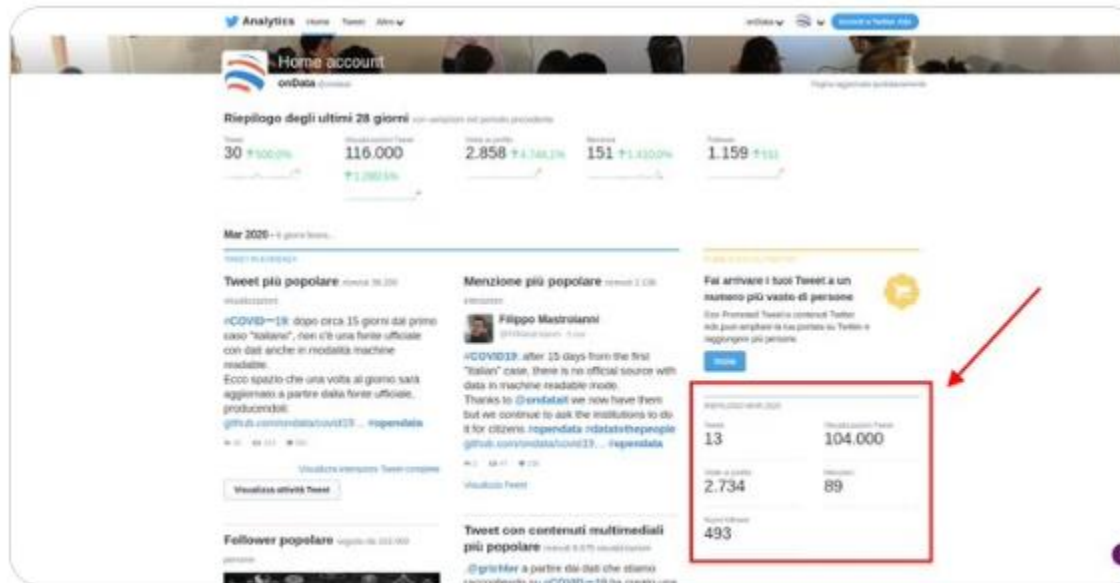
Affinché il lavoro del Team Open Data possa essere incisivo all'interno dell'amministrazione, è importante che tale team si **confronti** con il **livello più politico**, sia per ottenere da questo le necessarie "spinte", sia per offrire al decisore politico proposte e stimoli.



## #coronavirus

Dopo alcuni giorni di pressione da parte di migliaia di persone, @DPCgov ha pubblicato gli #opendata sulla diffusione del virus.

After a few days of strong pressure by thousands, @DPCgov published #opendata regarding the spread of #COVID-19 in Italy



# Organizzazione

## ***Responsabile per la transizione digitale***

*"SI RACCOMANDA di demandare al Responsabile per la transizione digitale (RTD) il compito di costituire un gruppo di lavoro dedicato al processo di apertura dei dati e all'implementazione delle presenti Linee Guida (n.d.r. [le nuove linee guida open data](#), attualmente in consultazione) all'interno dell'organizzazione dell'Ente. Il RTD deve essere comunque coinvolto in tutto il suddetto processo."*

# Organizzazione

## *Responsabile per la transizione digitale*

"La [Circolare n. 3](#) del 1° ottobre 2018 del Ministro per la pubblica amministrazione 2, al fine di garantire la piena operatività dell'Ufficio del RTD, **raccomanda** di prevedere, nell'atto di **conferimento** dell'**incarico** o di nomina, oltre che i compiti espressamente previsti, anche il potere, tra l'altro, di **costituire gruppi** tematici per singole attività e/o adempimenti come può essere il processo di **apertura e pubblicazione dei dati.**"

# Organizzazione

## *Responsabili e/o referenti tematici*

*"È bene prevedere un **responsabile** e il coinvolgimento dei **responsabili e/o referenti tematici** che gestiscono e trattano dati nell'ambito delle **singole unità organizzative**, oltre alle **figure** che possano fornire il necessario **supporto** per **l'analisi della qualità dei dati**, per la **definizione delle interfacce di accesso ai dati**, per la **promozione di applicazioni sviluppate a partire dai dati pubblicati**"*

# Organizzazione

*"È altresì importante garantire il **raccordo** e la consultazione con le altre **figure** coinvolte nel processo di **digitalizzazione** della pubblica amministrazione, ovvero **responsabile** per la **conservazione documentale**, **responsabile** per la **prevenzione della corruzione** e la **trasparenza**, **responsabile** per la **protezione dei dati personali**, **responsabile** dei sistemi **informativi** (se non coincide con il RTD), **responsabile** per la **sicurezza**"*

# Organizzazione

Incontrarsi:

- Internamente
  - Riunioni
  - Modalità di comunicazioni definite
  - Formazione
- Esternamente
  - Staff/progetto dedicato
  - Riunioni
  - Modalità di comunicazioni definite
  - hackathon/eventi



# Censimento

# Censimento

- Si parte da quelli esistenti, quelli "nativi";
- Il responsabile open data può lanciare ad esempio un questionario strutturato (formati, licenze, aggiornamento, referente, modalità accesso, ...);
- Riusare ciò che di prezioso già si raccoglie, come l'anagrafica delle "Basi di Dati", esito dell'[Art. 24-quater, comma 2, del D.L. n. 90/2014](#)

- > Presentazione
- > Staff
- > Rete di rilevazione
- > Dati
- > Previsioni
- > Difesa fitosanitaria
- > Progetti
- > Elaborazioni e mappe
- > Immagini da satellite
- > Pubblicazioni
- > AGROSERVIZI
- > Bollettini riepilogativi
- > Cartografia tematica
- > Qualità delle acque
- > Regione Siciliana
- > PSR Sicilia 2014-2020

Scarica l'app "Agrometeo SIAS" per Android



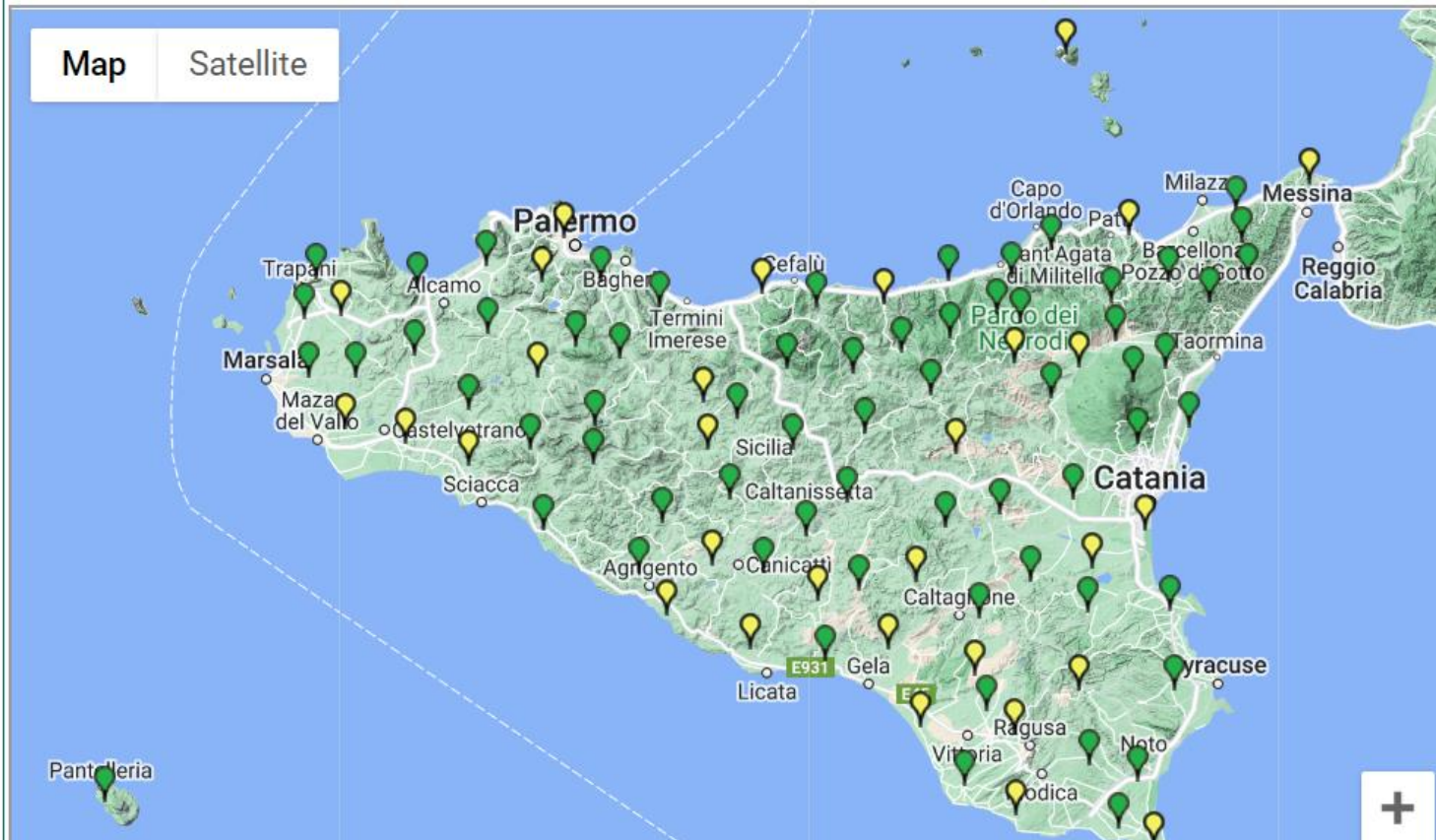
Grafici in tempo reale

Stazione v



Consultazione dati della rete di stazioni in telemisura

	P 1h	P 3h	P 12h	P 24h	P 48h	P 72h
Mappe dati in tempo reale (dati non validati)	Tinst	URinst	VV 2m max	VV 2m med	DV 2m	RG
Consultazione per grandezza meteorologica	Tutte	T min gior	T max gior	P gior	P 10 min (non validati)	
Metadati						



Pantelleria



## Cerca le basi di dati

Per titolo e descrizione



Per parola chiave

Per nome amministrazione

## Cerca per territorio

Regionale

Provinciale

Dataset Trovati: **1265**

### Personale

Il dataset contiene i dati anagrafici per la gestione dei dipendenti, servizio, assenze ed...

Publicato da: ISTITUTO SUPERIORE - IS 'MAJORANA'

Data di ultima modifica: 2021-04-02

ASSENZE

CONGEDI

CONTRATTI

SERVIZIO

lavoro

# Censimento

- Andare oltre gli obblighi, oltre il "nativo";
- Le informazioni che, una volta aperte, possano abilitare nuove forme di riutilizzo dell'informazione
  - La PA soggetto erogatore di servizi pubblici, ma anche espositore di patrimonio informativo che abilita nuove forme di business sul mercato;
- Si raccomanda anche un approccio "*demand-driven*", per individuare i dati nativi, che tenga conto dell'impatto economico e sociale nonché del livello di interesse degli utilizzatori suddivisi opportunamente per categorie (e.g., cittadini, imprese, altre pubbliche amministrazioni), dei loro requisiti e delle loro necessità.



# Censimento

I dati dalla "folla"





**TerremotoCentrotalia**  
TerremotoCentrotalia è un progetto non profit, organizzato interamente da attivisti. È nato per condividere informazioni utili e verificate sugli eventi sismici che hanno... leggi di più

Sostenitori **0** Dataset **5**

**ORGANIZZAZIONI**

TerremotoCentrotalia **5**

**GRUPPI**

*Non ci sono Gruppi che corrispondono a questa ricerca*

**TAG**

terremoto **5**

terremotocentroitalia **5**

centroitalia **1**

Home > Organizzazioni > TerremotoCentrotalia

Dataset **Flusso di attività** Informazioni

Cerca dataset...

## 5 dataset trovati

Ordina per: Rilevanza

### Segnalazioni Aggiornate da TerremotoCentrotalia

#### TerremotoCentrotalia

L'archivio delle segnalazioni di [www.terremotocentroitalia.info](http://www.terremotocentroitalia.info)

**VISUALIZZAZIONI** **RISORSE** 3 **FORMATO** CSV, JSON, GeoJSON

### Storie di Successo di TerremotoCentrotalia

#### TerremotoCentrotalia

Le storie da raccontare del progetto [terremotocentroitalia.info](http://terremotocentroitalia.info)

**VISUALIZZAZIONI** **RISORSE** 1 **FORMATO** CSV



data.europa.eu  
@EU\_opendata



We have just published a report on [#CitizenGeneratedData](#) 📖 Its main objective is to allow public administrations to better understand whether and how this kind of data may be made available in their [#OpenData](#) portals. Read it here: [europa.eu/!9kPCVW](https://europa.eu/!9kPCVW)

[Traduci il Tweet](#)

## Data.europa.eu and Citizen-generated Data

*Opportunities and challenges associated to the inclusion of citizen-generated data in data.europa.eu*

## 6. Conclusions, recommendations and future work

Based on this initial analysis, we derived the following conclusions.

- C1.** All portals had very few citizen-generated datasets, both in absolute terms and relative to the number of datasets hosted by the portals.
- C2.** In most cases, citizens are primarily involved in generating or collecting the data, but the remaining work required to publish the data is driven by public administrations. The efforts are initiated by public administrations rather than bottom-up by citizens, who are also less involved in curating or maintaining the data.
- C3.** The most frequent areas of CGD published in open data portals are: 'questions and answers', 'surveys' and 'statistics'. This is complementary to CS datasets (Ponti and Craglia, 2020).
- C4.** Primary CGD is more common than secondary CGD in open data portals.
- C5.** Most of CGD that is made available is shared with open licenses.
- C6.** Almost 30 % of the studied CGD datasets are available in open formats like JSON and XML. A much smaller percentage uses proprietary formats, typically XLSX.
- C7.** None of the studied portals included documentation about how to contribute and use CGD, nor about specific procedures to ensure data quality in this context. In fact, CGD is not explicitly identified as a data collection approach.
- C8.** In general terms – some exceptions are discussed in Section 5 – portals do not offer tools to facilitate citizen contributions, either at the level of datasets (upload their own data) or individual records (change, curate and maintain existing data).
- C9.** Following from **C2** and **C8**, we could find no evidence of participatory approaches to design data pipelines or collect and implement feedback from citizens on broader data strategy.
- C10.** No general guidelines on how to govern CGD in open data portals are provided; this seems to limit the emergence of more of these types of datasets.

# Organizzazione e Censimento



# Coinvolgimento

# Coinvolgimento

- Essere guidati dalla domanda
- Inserire dati nel contesto
- Supportare conversazioni intorno ai dati
- Creare capacità, competenze e reti
- Collaborare su dati come una risorsa comune

# Analisi giuridica



# Analisi giuridica

L'analisi giuridica delle fonti mira a valutare:

- Privacy
- Proprietà intellettuale della sorgente
- Licenza di rilascio
- Limiti alla pubblicazione
- Segretezza
- Temporalizzazione
- Trasparenza



Privacy	hai considerato il rischio di de-anonimizzazione del tuo dataset prima di pubblicarlo?
Privacy	esponi dei servizi di ricerca tali da poter filtrare i dati in modo da ottenere un solo record geolocalizzato, che sia facilmente riconducibile ad una persona fisica?
Proprietà intellettuale della sorgente	il dataset è stato creato da uno o più dipendenti della tua pubblica amministrazione nell'ambito della loro attività lavorativa? I singoli elementi del dataset suscettibili di autonoma protezione (es., immagini, fotografie, testi in qualche modo creativi) sono stati a loro volta prodotti da uno o più dipendenti della tua pubblica amministrazione nell'ambito della loro attività lavorativa?
Proprietà intellettuale della sorgente	l'amministrazione è proprietaria dei dati, anche se non sono stati creati direttamente da suoi dipendenti??
Proprietà intellettuale della sorgente	sei sicuro di non usare dati per i quali vi è una licenza o un brevetto di terzi?
Proprietà intellettuale della sorgente	se i dati non sono della tua amministrazione hai un accordo o una licenza che ti autorizzi a pubblicarli?
Licenza di rilascio	stai rilasciando i dati di cui possiedi la proprietà accompagnati da una licenza?
Licenza di rilascio	hai incluso anche la clausola di salvaguardia «Questo dataset contiene informazioni indirettamente riferibili a persone fisiche. In ogni caso, i dati non possono essere utilizzati al fine di identificare nuovamente gli interessati.»?
Limiti alla pubblicazione	hai verificato che non vi siano impedimenti di legge o contrattuali che per la pubblicazione dei dati?

# Analisi della qualità dei dati

# Analisi qualità dei dati

- **accuratezza** (sintattica e semantica) - il dato, e i suoi attributi, rappresenta correttamente il valore reale del concetto o evento cui si riferisce;
- **coerenza** - il dato, e i suoi attributi, non presenta contraddittorietà rispetto ad altri dati del contesto d'uso dell'amministrazione titolare;
- **completezza** - il dato risulta esaustivo per tutti i suoi valori attesi e rispetto alle entità relative (fonti) che concorrono alla definizione del procedimento;
- **attualità** (o tempestività di aggiornamento) - il dato, e i suoi attributi, è del "giusto tempo" (è aggiornato) rispetto al procedimento cui si riferisce.

# La "bonifica"

# Bonifica

- È un'occasione importante, perché può portare al sistema di origine, gli esiti;
- Può essere **basata sui dati**;
- Può essere (preferibile) **basata sui processi**.

Individuare le persone  
che "producono" i dati



# Le persone

- Chi è responsabile?
- Chi è "titolare"?
- Chi aggiunge elementi informativi?

*"Quello che accade sovente nelle amministrazioni è che i dati sono gestiti da singoli funzionari, nell'ambito di processi "verticali" chiusi a livello di dipartimento e molto spesso ancorati alle conoscenze di una persona specifica. Accade così che elementi conoscitivi importanti siano delocalizzati tra i servizi di competenza, senza che tuttavia sussista una gestione federata e complessiva della risorsa dati. "*

# I "Panieri"

# I "Panieri"

La disponibilità di banche dati pubbliche in formato aperto rappresenta l'**indicatore di risultato** (2RIS) di riferimento nell'**Accordo di Partenariato 2014-2020** tra la Commissione Europea e l'Italia, ed è uno degli elementi volti a esprimere lo stato del processo di **valorizzazione dei dati in formato aperto** e dell'implementazione della direttiva PSI (*Public Sector Information*) in Italia.

Il Paniere dinamico di dataset risponde quindi alla duplice esigenza di dare **contenuto quantitativo** e **qualitativo** all'**indicatore 2RIS** sulla **Disponibilità di banche dati pubbliche in formato aperto**.

Dataset (nazionali e regionali) riferibili agli **impegni di governo**, alle **iniziative** delle **amministrazioni**, alle **richieste** degli **stakeholders**, agli impegni assunti nel contesto dell'**Open Government Partnership**, alla strategia di riferimento delineata con il **Piano triennale ICT**.

# Il "Paniere" nazionale

Scheda Dimissioni ospedaliera	Assistenza in emergenza sanitaria territoriale	Caserme vuote/dismesse
Accesso ai luoghi della cultura	Attività delle strutture di ricovero	Catalogo Beni e Servizi
Agrometeo	Aziende confiscate alle mafie	Classificazione Ateco
Amministrazioni del Programma Razionalizzazione degli Acquisti	Aziende Partecipate	Classificazione delle professioni
Anagrafe edifici pubblici e a uso pubblico [2]	Banda Larga e Ultra Larga	Classificazione internazionale della spesa pubblica (COFOG)
Anagrafe Imprese	Bandi e Gare	Consumi di dispositivi medici
Anagrafica delle strutture scolastiche	Beni confiscati e sequestrati alle mafie	Contratti pubblici
Analisi economica dei costi degli infortuni e delle malattie professionali	Bilanci Pubblici	Database GeoTopografico [4]
Appalti innovativi	Bollettino Agrit	Dati elettorali
Archivio Nazionale Numeri Civici e Strade Urbane	CAP	Denunce malattie professionali

Dimissioni ospedaliere	Mappa Nazionale: DBPrior	Piano paesaggistico regionale [3]
Efficientamento energetico	Marchi di qualità e disciplinari	Registro Imprese
Eventi e luoghi della cultura	Normattiva	Repertorio nazionale per gli interventi per la Difesa del Suolo
Finanza delle strutture scolastiche	Open Coesione	Rete mareografica e dati meteorologici associati
Giornale di bordo (pesca)	Open Coesione Plus	Rete Ondametrica Nazionale
Indicatori di popolazione e abitazioni	Sostanze inquinanti per l'ambiente	Rifiuti urbani e raccolta differenziata
Infrastrutture e mobilità	Specie marine aliene	Schede di catalogo
Istituti Penitenziari	OpenCUP	Sicurezza strutture scolastiche
Libri scolastici	Partecipazioni	SIOPE
Malattie Professionali	Piani di Assetto Idrogeologico / Piani di Gestione del Rischio Alluvioni dei Distretti Idrografici [2]	Sistemi locali di lavoro

# Il "Paniere" regionale



Agricoltura biologica	Catasto impianti termici	Elenco delle Discariche
Agriturismi (a)	Comunicazioni obbligatorie	Elenco impianti sportivi
Aree a rischio incendi boschivi	Contributi e sussidi	Elenco organizzazioni non governative (ONG)
Aree protette e Rete Natura 2000	Cooperative sociali	Elenco sensori meteo
Aziende esercenti servizi di trasporto pubblico di linea	Dati sensori meteo	Energia
Beni demaniali regionali	ERP - Edilizia residenziale pubblica	Epidemiologia - Cause di morte
Bonifiche siti contaminati	Edicole	Esercizi di erogazione carburante
Carta degli Incendi	Elenchi strutture commerciali	Farmacie (dataset Min. della Salute)
Carta tecnica regionale numerica (CTRN)	Elenco RSA	Fattorie didattiche
Catasto Cave	Elenco comunità per minori	Incidenti stradali (a)

Istituti pubblici di assistenza e beneficenza (IPAB)	Polizie locali
Limiti amministrativi dei comuni	Presenze turistiche
Locazioni passive	Prezzario dei Lavori Pubblici della regione
Operatori accreditati ai servizi al lavoro	Qualità dell'aria
Orari trasporto pubblico regionale (a)	Rete Viaria
Pagamenti	Sale cinematografiche attive
Parafarmacie (dataset Min. della Salute)	Strutture ricettive (a)
Patrimonio Indisponibile	Terminal intermodali
Percorsi di Istruzione e Formazione Tecnica Superiore (IFTS)	
Percorsi di formazione superiore (ITS)	

E i comuni??

# I Comuni

- Sono l'interfaccia più diffusa, ricca e vicina alla persone;
- Sono diversissimi tra loro, ma hanno tanto in Comune 😊;
- Ci sono due progetti interessantissimi legati agli open data dei Comuni
  - Regione Lombardia
  - Regione Piemonte

## Trasformazione Digitale in Lombardia

Trasformazione Digitale in Lombardia / Open Data: pubblicato lo standard dei 50 dataset del paniere regionale

Home

Trasformazione digitale in Lombardia

Analisi avanzate dei dati

Pagamenti verso la PA

API e interoperabilità

Procedimenti e CDS  
TELEMATICA

**Open Data**

Chi usa gli open data di Regione Lombardia

Formazione su open data

Centri di competenza

### Open Data: pubblicato lo standard dei 50 dataset del paniere regionale



Data avviso: 09 agosto 2018

Il **'paniere' regionale di dataset** per gli enti locali è un elenco di 50 dataset, appartenenti a 10 categorie o aree tematiche, selezionati tra quelli che si sono rivelati più utili anche sulla base del monitoraggio degli open data messi a disposizione da altre amministrazioni italiane, e rappresenta un elenco di dataset che tutti i Comuni possono rendere pubblici.

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.

Il 'paniere' è frutto di un processo partecipativo che ha coinvolto gli enti del territorio che hanno aderito a un bando per l'erogazione di contributi economici in compartecipazione alle attività di adeguamento dei

Non si tratta solo di un elenco di 50 “titoli”, ma per ognuno è stato anche proposto un **tracciato record di base**

- 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.

# Tracciato record

- Quali sono le colonne, i campi, che definiscono la mia risorsa dati?
- Qual è il titolo e la loro descrizione?
- Che tipo di contenuto contiene ciascuna colonna? Testo? Date? Numeri decimali? ecc., ecc..
- È obbligatorio compilare quella colonna?
- A quale vocabolario controllato si fa riferimento per i valori di una determinata colonna?



## #35 Quantità rifiuti prodotta

**Descrizione:** Quantità di rifiuti prodotti dalla raccolta differenziata e non differenziata, specificandone categoria e modalità di raccolta.

**Tipologia:** Dataset tabellare

**Frequenza minima:** Mensile

Denominazione campo	Descrizione	Tipo	Obbligo
<b>Anno</b>	Anno di riferimento	Numero	M
<b>Macro_Categoria_Rifiuti</b>	Indicare se si tratta di rifiuti non differenziati, raccolta differenziata o Inerti e rifiuti da costruzione e demolizione	Testo	M
<b>Categoria_Rifiuti</b>	Indicare la categoria di rifiuti (es. carta e cartone, vetro, ecc.)	Testo	M
<b>ID_Categoria_Rifiuti</b>	Codice della Categoria rifiuti	Testo	O
<b>Modalità_Raccolta</b>	Indicare la modalità di raccolta (es. meccanizzata, porta a porta, area attrezzata)	Testo	M
<b>Periodo</b>	Indicare il mese di riferimento in numero	Numero	M
<b>Quantità</b>	Indicare la quantità in kg	Numero	M

Tabella 5. Tracciato standard Quantità rifiuti prodotta



## Standard dei dataset del ‘paniere’ regionale per gli enti locali

### PREMESSA

Regione Piemonte, riutilizzando l’ottimo lavoro fatto dalla [Regione Lombardia](#), e nell’ambito delle iniziative per la costituzione dell’ecosistema piemontese dei dati aperti, intende con questo documento pubblicare l’elenco dei dataset e il relativo tracciato che gli Enti Locali sono invitati a pubblicare in Open Data. La pubblicazione dei dati in open data da parte degli Enti Locali avviene seguendo quanto indicato nelle linee guida Regionali<sup>1</sup>, e quanto da loro pubblicato è visibile sul portale regionale dei dati aperti della Regione Piemonte [www.dati.piemonte.it](http://www.dati.piemonte.it).

Per il caricamento dei dati gli Enti Locali hanno a disposizione la [Piattaforma Yucca – Smartdataplatform](#), strumento messo a disposizione dalla Regione Piemonte a tutti gli Enti del Territorio per il caricamento e l’elaborazione dei propri dati. Per l’accesso, i termini d’uso e le funzionalità del back-office di Yucca, si faccia riferimento alla documentazione disponibile all’indirizzo <https://www.dati.piemonte.it/#/aderire>.

<b>Tema</b>	<b>Ambito tematico YUCCA</b>	<b>Sotto ambito tematico YUCCA</b>	<b>Nome Dataset</b>	<b>Numero</b>
AMBIENTE	Ambiente	Ambiente urbano	Aree Verdi	<a href="#">#17 Aree Verdi</a>
AMBIENTE	Ambiente	Ambiente urbano	Aree Verdi Informazioni	<a href="#">#18 Aree Verdi Informazioni</a>
AMBIENTE	Ambiente	Ambiente	Carta degli incendi	<a href="#">#1 Carta degli incendi</a>
AMBIENTE	Ambiente	Rifiuti	Elenco degli impianti di depurazione	<a href="#">#34 Elenco degli impianti di depurazione</a>
COMMERCIO	Commercio	Servizi	Pratiche SUAP	<a href="#">#20 Pratiche SUAP</a>
COMMERCIO	Commercio	Servizi	Pubblici esercizi	<a href="#">#47 Pubblici esercizi</a>
COMMERCIO	Commercio	Servizi	Servizi alla persona	<a href="#">#45 Servizi alla persona</a>
CULTURA	Cultura	Cultura	Beni artistici e architettonici	<a href="#">#50 Beni artistici e architettonici</a>
CULTURA	Cultura	Comunicazione	Eventi	<a href="#">#48 Eventi</a>
ENERGIA	Energia	Energia	Albo provinciale manutentori impianti termici	<a href="#">#30 Albo provinciale manutentori impianti termici</a>



## INDICAZIONI PRELIMINARI

### Formato dei dati

I dati che possono essere caricati mediante l'uso della piattaforma YUCCA, possono essere in formato binary (nel caso di filmati, immagini o file binari da allegare) oppure in formato testo csv.

Le colonne del dataset derivante da file .csv possono essere di tipo:

- Numero (intero, intero lungo, decimale)
- Testo
- Booleano
- Data
- Coppia di coordinate geografiche (latitudine e longitudine)

Per ogni colonna è possibile specificare l'unità di misura (esempio kWh se il dato è di tipo energetico o % se il dato è una frazione).

# I metadati (catalogo)

# Linee guida per i cataloghi dati

- Le trovate pubblicate su Docs Italia (lo spazio dei Documenti pubblici, digitali);
- Forniscono indicazioni su
  - i **temi** dei dati
  - i **vocabolari controllati** da utilizzare
  - nonché una descrizione degli **elementi** principali del **profilo di metadazione** con le relative proprietà
- Il profilo di metadazione si chiama **DCAT-AP\_IT**





# Il luogo per la divulgazione e la consultazione dei documenti pubblici digitali.

 Search all the docs

## Documenti in evidenza



PIANO TRIENNALE PER  
L'INFORMATICA NELLA  
PUBBLICA  
AMMINISTRAZIONE

### [Codice dell'amministrazione digitale](#)

Lo Stato, le Regioni e le autonomie locali assicurano la disponibilità, la gestione, l'accesso, la trasmissione, la conservazione e la fruibilità dell'informazione in modalità digitale e si organizzano ed



PIANO TRIENNALE PER  
L'INFORMATICA NELLA  
PUBBLICA  
AMMINISTRAZIONE

### [Piano triennale ICT](#)

Il Piano triennale è costruito sulla base di un Modello strategico di evoluzione del sistema informativo della Pubblica amministrazione e indirizza il piano delle gare, il piano dei finanziamenti e i piani triennali delle singole PA.



DESIGN

### [Linee guida di design per i servizi web della PA](#)

Linee guida di design per i servizi web della PA



CLOUD ITALIA

### [Strategia Cloud Italia](#)

Documento sintetico di indirizzo strategico per l'implementazione e il controllo del Cloud della PA

# Descrivere un catalogo

- **Titolo** > dct:title
- **Descrizione** > dct:description
- **Editore** > dct:publisher
- **Data ultima modifica** > dct:modified
- **Dataset** > dcat:dataset

# Descrivere un dataset

- **Identificativo del dataset** > dct:identifier
- **Titolo del dataset** > dct:title
- **Descrizione del dataset** > dct:description
- **Ultima modifica del dataset** > dct:modified
- **Temi del dataset** > dct:theme
- **Titolare del dataset** > dct:rightsHolder
- **Frequenza aggiornamento dataset** > dct:accrualPeriodicity
- **Distribuzione del dataset** > dct:distribution

# Temi dei dati

1. Agricoltura, pesca, silvicoltura e prodotti alimentari
2. Economia e Finanze
3. Istruzione, cultura e sport
4. Energia
5. Ambiente
6. Governo e settore pubblico
7. Salute
8. Tematiche internazionali
9. Giustizia, sistema giuridico e sicurezza pubblica
10. Regioni e città
11. Popolazione e società
12. Scienza e tecnologia
13. Trasporti

# Descrivere risorsa/distribuzione

- **Formato Distribuzione** > dct:format
- **URL di accesso** > dcat:accessURL
- **Licenza** > dct:license

# I metadati sui contenuti

Purtroppo avviene troppo di rado

# Dizionari di dati


- Il **dizionario dei dati** è un complemento essenziale per qualsiasi serie di dati, poiché **fornisce** all'utente le **informazioni** sufficienti per elaborare e comprenderne il contenuto senza ambiguità.
- Il suo scopo è quello di **garantire** che la **struttura del set di dati** sia **definita in termini** facilmente **comprensibili** dagli utenti.
- **Ogni set di dati** pubblicabile deve **includere** il suo **dizionario di dati** come documento separato, di solito accessibile tramite un URL presente nello spazio di download dei dati.



# Dizionari di dati

- Il contenuto di un **dizionario di dati** può essere espresso in diversi modi, anche **come un file di testo** che **descrive il contenuto di ogni colonna** del set di dati.
- Le **caratteristiche o annotazioni** espresse nel **dizionario dei dati** sono le **proprietà** delle tabelle, **colonne, righe** e **celle** che compongono l'insieme di dati.
- Quello che segue è un esempio di un dizionario di dati espresso come file di testo, che può essere fornito attraverso un server web. Per esempio, tramite l'URL:  
[https://sito\\_di\\_esempio.it/automobili.csv-metadata.txt](https://sito_di_esempio.it/automobili.csv-metadata.txt).

## automobili.csv-metadata.txt

File di dati: [https://sito\\_di\\_esempio.it/automobili.csv](https://sito_di_esempio.it/automobili.csv) 

Descrizione: tabella con dati sulle auto d'epoca

Editore: Autore di esempio

Colonna 1:

Titolo: marca

Descrizione: Questo campo contiene informazioni sulla marca e il modello di c

Tipo di dati: stringa

Colonna 2:

Titolo: anno

Descrizione: Questo campo contiene informazioni sull'anno di fabbricazione di

Tipo di dati: data

Colonna 3:

Titolo: cilindri

Descrizione: Questo campo contiene informazioni sul numero di cilindri di ogn

Tipo di dati: intero

Colonna 4:

Titolo: consumo

Descrizione: Questo campo contiene informazioni sul consumo medio di ogni ve:

Tipo di dati: decimale

Colonna 5:

Titolo: potenza

Descrizione: Questo campo contiene informazioni sulla potenza di ogni veicolo

Tipo di dati: decimale

Colonna 6:

Titolo: accelerazione

Descrizione: Questo campo contiene dati sull'accelerazione di ogni veicolo, r

Tipo di dati: decimale

Nome campo	Descrizione	Description	Formato	Esempio
<b>data</b>	Data dell'informazione	Date of notification	YYYY-MM-DDTHH:MM:SS (ISO 8601) Ora italiana	2020-03-05T12:15:45
<b>stato</b>	Stato di riferimento	Country of reference	XYZ (ISO 3166-1 alpha-3)	ITA
<b>codice_regione</b>	Codice della Regione (ISTAT 2019)	Code of the Region (ISTAT 2019)	Numero	13
<b>denominazione_regione</b>	Denominazione della Regione	Name of the Region	Testo	Abruzzo
<b>lat</b>	Latitudine	Latitude	WGS84	42.6589177
<b>long</b>	Longitudine	Longitude	WGS84	13.70439971
<b>ricoverati_con_sintomi</b>	Ricoverati con sintomi	Hospitalised patients with symptoms	Numero	3
<b>terapia_intensiva</b>	Ricoverati in terapia intensiva	Intensive Care	Numero	3
<b>totale_ospedalizzati</b>	Totale ospedalizzati	Total hospitalised patients	Numero	3
<b>isolamento_domiciliare</b>	Persone in isolamento domiciliare	Home confinement	Numero	3
<b>totale_positivi</b>	Totale attualmente positivi (ospedalizzati + isolamento domiciliare)	Total amount of current positive cases (Hospitalised patients + Home confinement)	Numero	3

```
schema:
  - missingValues:
    - "NA"
    - "NaN"
    - null
    - ""
  - primaryKey:
    - id
  - fields:
    - type: integer
      name: id
    - type: string
      name: nome
    - type: date
      name: data-accesso
      constraints:
        - maximum: '2022-05-15'
      description: La data di accesso dell'utente nel sistema
      example: '2020-05-15'
    - type: integer
      name: eta
    - type: string
      name: ruolo
      constraints:
        - enum:
            - A
            - B
    - type: string
      name: email
      format: email
      required: true
    - type: string
      name: istat
      constraints:
        - pattern: ^[0-9]{6}$
    - type: geopoint
      name: geo
```

Frictionless Standard



# LADIDATTICA

FORMAZIONE ONLINE PER LA PA

## OPEN DATA NELLA PA

LA PREPARAZIONE DEI DATASET E DELLE RISORSE

# GRAZIE PER L'ATTENZIONE

**Andrea Borruso**

Esperto di dati aperti



Formez PA

