

OPEN DATA ED ENERGIA, QUALI SOLUZIONI ADOTTARE

I DATI APERTI SONO INDISPENSABILI PER L'ELABORAZIONE DI POLITICHE E STRATEGIE. IL LORO USO HA GRANDI POTENZIALITÀ NON SOLO PER GLI UTENTI ESTERNI, MA ANCHE PER GLI ENTI CHE LI PRODUCONO. L'APPROCCIO OPEN DATA RICHIEDE REGOLE ORGANIZZATIVE E TECNOLOGICHE PER GARANTIRE COMPLETEZZA, AGGIORNAMENTO E SICUREZZA.

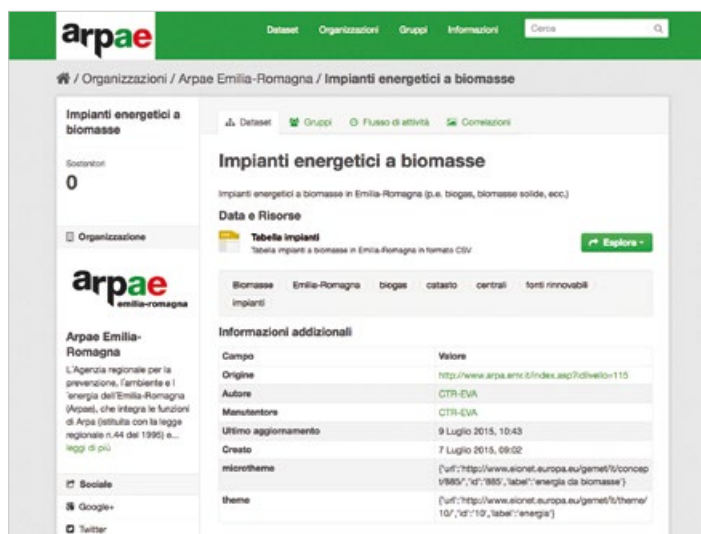
I dati aperti risultano indispensabili per le analisi sull'energia: gli innumerevoli indicatori di interesse per l'elaborazione di politiche e strategie derivano da analisi e simulazioni su dati interdisciplinari, spesso provenienti da fonti totalmente indipendenti e apparentemente scorrelate. Si pensi ai dati su materiali da costruzione, consumi energetici, caldaie a combustione o teleriscaldamento. I dati territoriali su questi comparti sono difficoltosi, se non impossibili, da reperire in rete. Risulta perciò necessario identificare pazientemente le fonti qualificate e interagire con interlocutori in grado di capire la domanda e fornire le risposte (se disponibili).

Più semplice è reperire i dati meteorologici (fondamentali per il calcolo del bilancio energetico): le strutture preposte stanno già affrontando il tema dei *big data*, e la predisposizione di previsioni su griglia in tempo reale sfrutta tecnologie già in grado di disseminare un'enorme mole di dati.

La complessità della materia, con gli sforzi tecnologici che comporta la fornitura di *open data*, sta portando a nuovi paradigmi, il principale dei quali riguarda il fatto che i dati aperti servono sì alla società civile ma, anzitutto, hanno un potenziale enorme per gli usi interni alle stesse amministrazioni/aziende che li producono. Sappiamo, ad esempio, quanto è difficile mantenere l'aggiornamento dei dati e la relativa tracciabilità, o effettuare correzioni una volta che i dati sono stati rilasciati. Questi aspetti, che riguardano il ciclo di vita dei dati, devono essere gestiti non solo per i fruitori esterni ma, anzitutto, per chi li produce e ha l'onere della manutenzione. La formalizzazione del ciclo di vita dei dati risulta fondamentale per i fruitori interni alle amministrazioni/aziende, proprio per produrre in modo efficiente la reportistica, sia richiesta

FIG. 1
OPEN DATA

Il portale open data di Arpae Emilia-Romagna (dati.arpae.it).



a livello istituzionale, sia per la società civile. Quindi "la pubblicazione di dati non deve essere un'attività a latere, cioè accessoria, ma compenetrare l'attività istituzionale. Chi produce il dato lo dovrebbe depositare in un contenitore che, in un processo di integrazione e omogeneizzazione, ed entro i limiti della normativa sulla privacy, diventa immediatamente accessibile sia agli *stakeholder* interni (a fini di utilizzo istituzionale), sia contemporaneamente a quelli esterni. L'effetto immediato è che non si generano né sovrapposizioni, né dispersione di risorse." (Cattani e Fustini, *Ecoscienza*, 2/2015).

Poiché i dati necessari ai piani energetici e al relativo monitoraggio richiedono continui aggiornamenti in termini di completezza e grado di approfondimento, servono strumenti di pubblicazione che ne favoriscano la gestione in termini di sicurezza informatica, guardando anche al livello organizzativo: si tratta di definire un "workflow multireferenziale" che consenta (al personale che ne ha titolo) di inserire dati nel database al quale si avrà accesso in fase di pubblicazione dei dati aperti (Cattani e Rossi, *Ecoscienza*, 5/2015).

Si tratta di utilizzare al meglio le potenzialità dell'*approccio open data* (regole chiare per titolarità, riutilizzo ecc., specifiche tecnologiche, indicizzazione e metadati, web semantico ecc. e anche, più in generale, una propensione alla condivisione "nativa" dei dati) in una dimensione interna nella stessa organizzazione o tra organizzazioni; questo viene definito "*open data interno*", ponendo eventualmente ulteriore attenzione sulle regole di condivisione dei dati permettendo, potenzialmente, di rendere disponibili dati altrimenti non liberabili (condivisibili) in generale. Quindi guardiamo in termini di *open data interno*: gli utenti "fornitori" hanno a disposizione uno strumento di inserimento dati che consente modifiche e verifica di tutti i dati inseriti, mantenendo le titolarità e la storicizzazione delle modifiche. Gli stessi utenti, insieme ad altri fruitori interni, utilizzano questa base dati come quella principale e istituzionale per la reportistica, lasciando che i dati presenti nelle procedure amministrative (possibili sorgenti "grezze") siano eventualmente da modificare *ex-post*. Per fare questo si torna al livello organizzativo: servono regole chiare

e certe per la condivisione dei dati, sapendo che i dati derivati, aggregati e normalizzati saranno a supporto delle azioni e delle politiche di gestione.

La produzione di queste “nuove” basi dati idonee alla pubblicazione di dati aperti sembrerebbe dover coinvolgere pesantemente i sistemi informativi aziendali. Questo aspetto può risultare una criticità, poiché nel mandato dei sistemi informativi generalmente non è previsto il ripensamento dei processi e, quindi, può risultare molto difficoltoso il reperimento di risorse da dedicare a questi temi.

Per fortuna, oggi ci viene in aiuto la tecnologia basata sul *public cloud*: infrastrutture già a misura di utente finale che incorporano tutti i necessari criteri di sicurezza in termini di:

1. disponibilità in ogni momento da parte di chi ha diritti di accesso
2. integrità, ovvero immutabilità da parte chi non ne ha diritto
3. confidenzialità, ovvero accesso consentito esclusivamente da chi ne possiede i diritti.

Si noti che il foglio elettronico residente su Pc (o relativo file csv) non risponde a nessuno di questi criteri.

A titolo di esempio, immaginiamo una struttura che già utilizza fogli elettronici come contenitori di dati idonei alla reportistica e alla pubblicazione. Il foglio non è un *database*: non può tenere traccia delle modifiche, non è multiutente e i relativi dati non possono essere pubblicati in modo sincrono.

Uno degli strumenti idonei per produrre semplicemente dati aperti pubblicabili rimuovendo le criticità evidenziate è il foglio elettronico disponibile nella suite Google Drive.

Si può partire da un classico foglio Excel per il caricamento dei dati attualmente disponibili (in *figura 1*, un esempio sui dati degli impianti a biomassa).

Come si può gestire un foglio Excel in un Pc e, contemporaneamente, ragionare in termini di *open data*? Un file xls non è *open data*, e non può nemmeno essere reso disponibile in rete in modalità *machine2machine*.

E poi come è possibile gestire più contributori? E la storia delle modifiche effettuate?

Per ovviare a tutto ciò, il foglio deve essere portato nativamente in una logica *open data* compatibile: si diceva di portarlo in *public cloud*.

Una volta che il file è “convertito” all’interno del *public cloud*, si apre uno scenario di lavoro completamente nuovo, ovvero si possono definire i criteri di

FIG. 2
OPEN DATA

Esempio di mappa con gli impianti a biomassa in Emilia-Romagna presenti nel database di Arpae Emilia-Romagna.

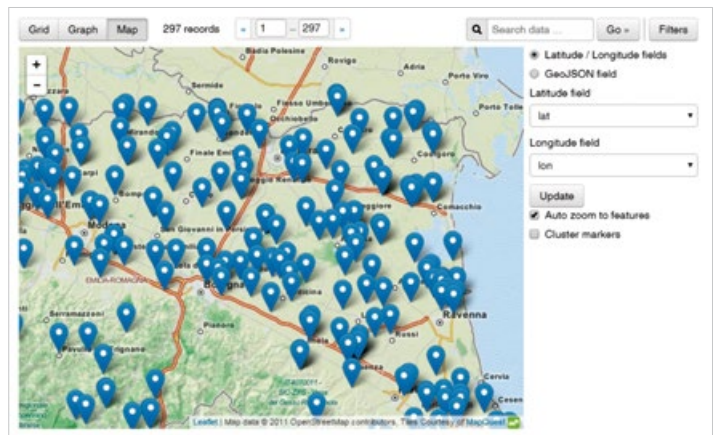


FIG. 3
OPEN DATA

Scheda di un impianto presente nel database di Arpae Emilia-Romagna.



accesso multiutente, dando privilegi di scrittura e lettura come se si trattasse di un applicativo basato su *database* e account con privilegi, ma è sempre lo stesso foglio elettronico.

Inoltre, lo strumento possiede intrinsecamente la storicizzazione di tutte le modifiche effettuate, senza possibilità di errore.

Quindi, ora che è accessibile in *cloud*, possiamo pubblicarlo in *open data*? No. In quel foglio ci possono essere dati soggetti a vincoli di *privacy*. E poi lo vogliamo tenere a uso “interno”, con eventuali note di lavoro.

Quello che invece pubblicheremo come *open data* sarà un nuovo foglio collegato “a caldo” (cioè in cui le modifiche effettuate sul foglio originale vengono automaticamente ereditate dal foglio pubblicato) con quello originale, ma saranno collegate solo le informazioni che vogliamo pubblicare. Ad esempio, potremo togliere le ragioni sociali ed evidenziare le coordinate geografiche.

Il nuovo foglio, collegato a caldo con quello di lavoro, potrà essere pubblicato direttamente e aggiornato in automatico da quello di lavoro.

A questo punto, sul portale aziendale *open data* (Ckan) il link da fornire sarà quello del foglio pubblicato e il meccanismo di pubblicazione potrà considerarsi terminato, con la rispondenza completa e puntuale dei criteri di sicurezza informatica declinati sopra.

Per concludere, il foglio Excel che prima stava relegato in un Pc ora è una fonte dati aggiornabile, sempre allineata e con un *end point* sempre raggiungibile. Anche dal personale interno. Pensare in termini di *open data* serve a migliorare l’organizzazione del lavoro.

Stefano Cattani¹, Massimo Fustini²

1. Arpae Emilia-Romagna
2. Regione Emilia-Romagna