

IL CICLO DI VITA DEGLI OPEN DATA AMBIENTALI

GLI OPEN DATA RICHIEDONO UN RIPENSAMENTO DEL PROCESSO DI PRODUZIONE E RILASCIO DEI DATI, IN PARTICOLARE PER QUANTO RIGUARDA IL FLUSSO DI VALIDAZIONE E PUBBLICAZIONE. VA INCLUSO IL FEEDBACK PER INTEGRARE E MIGLIORARE IL CONTENUTO INFORMATIVO, RENDENDO IL PROCESSO NON PIÙ UNIDIREZIONALE, MA CIRCOLARE.

Dopo aver presentato un *“Progetto per gli open data ambientali dell’Emilia-Romagna”* (Cattani-Fustini, in *Ecoscienza* 2/2015), cerchiamo di analizzare alcune possibili modalità di gestione operativa. Anzitutto è necessario un ripensamento del processo di produzione e rilascio dei dati, in particolare per quanto riguarda il flusso cosiddetto “di validazione” e la conseguente pubblicazione. Gli open data richiedono tempestività e capacità di gestire il *feedback* dai vari portatori di interesse, ma finora questi aspetti non sono mai stati considerati prioritari, a favore del processo di validazione o certificazione. L’attuale sistema di validazione presuppone che il produttore del dato possa migliorarne nel tempo la qualità, ma sempre sulla base esclusiva delle conoscenze interne. Di conseguenza, i sistemi informatici sono stati costruiti in modo autoreferenziale, senza la possibilità di gestire un vero *workflow* multireferenziale di gestione del patrimonio informativo collettivo. A questo si aggiunge il fatto che i dati, spesso, vengono ricavati da processi amministrativi con diversi responsabili di procedimento che non sono contemporaneamente referenti della pubblicazione, creando ulteriori rigidità.

Dobbiamo ripensare il ciclo di vita del dato finalizzato alla pubblicazione, inteso come processo che ne prevede la genesi e diversi trattamenti successivi (tra i quali una o più “validazioni”). Ma anche, e soprattutto, dobbiamo includere il *feedback* per integrare e migliorare il contenuto informativo, ovvero correggere i *dataset*. A oggi non esistono linee guida in materia, ma solo *best practices* maturate sulla base dell’organizzazione delle strutture coinvolte e, più spesso, della buona volontà dei singoli. Affrontiamo, a titolo di esempio, la gestione dei dati di un catasto, costituito da dati puntuali georeferenziati, ai quali vengono associati dati tabellari (attributi).

In genere un catasto ha tempi di aggiornamento nell’ordine delle settimane o dei mesi: le fonti dati possono essere disparate, e la titolarità dei dati può essere distribuita su più soggetti. A livello di pubblicazione, questa tipologia di dati è generalmente appoggiata su fogli elettronici, o semplici *database client*, o strati cartografici. Il flusso di produzione è unidirezionale: da diversi software di gestione vengono

“estratti” i dati di interesse, resi coerenti e pubblicati. Questo meccanismo non permette di correggere *ex-post*: un dato errato rimarrà tale nei software gestionali che lo hanno prodotto, poiché risulterà troppo complesso ripercorrere a ritroso il flusso. Soprattutto, si continuerà a ritenere che il dato “corretto” sia quello all’interno dei sistemi gestionali. Il primo cambio di paradigma è il seguente: si deve allocare il catasto

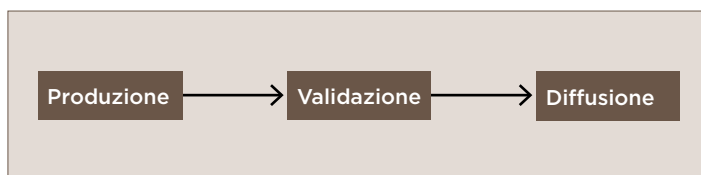


FIG. 1
PROCESSO LINEARE

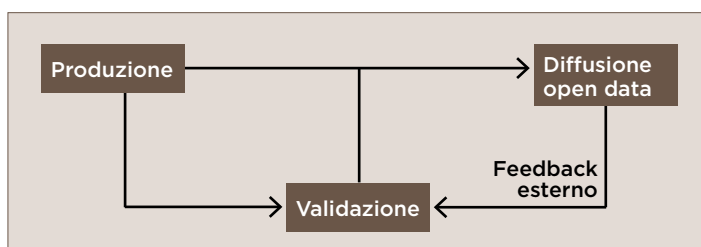


FIG. 2
PROCESSO CIRCOLARE

Impianti energetici a biomasse

Impianti energetici a biomasse in Emilia-Romagna (p.e. biogas, biomasse solide, ecc.)

Data e Risorse

Tabella impianti
Tabella impianti a biomasse in Emilia-Romagna in formato CSV Esplora

Biomasse Emilia-Romagna biogas catasto centrali fonti rinnovabili impianti

Informazioni aggiuntive

Campo	Valore
Origine	http://www.arpa.emr.it/index.asp?idlivello=115
Autore	CTR-EVA
Manutentore	CTR-EVA
Stato	active
Ultimo aggiornamento	9 Luglio 2015, 10:43
Creato	7 Luglio 2015, 09:02
microtheme	{'url':'http://www.eionet.europa.eu/gemet/it/concept/885','id':'885','label':'energia da biomasse'}
theme	{'url':'http://www.eionet.europa.eu/gemet/it/theme/10','id':'10','label':'energia'}

FIG. 3
ESEMPIO DI METADATI

informatico al di fuori dei sistemi gestionali, si attribuisce una responsabilità di manutenzione e aggiornamento sempre al di fuori dei sistemi gestionali e, solo dopo, l'organizzazione decide come mantenere allineati i due *dataset*. Il vantaggio di questa visione è quello di avere due diversi punti di vista: da una parte la fornitura del dato con i requisiti di tempestività di pubblicazione e rapidità di aggiornamento, dall'altra il processo di gestione amministrativa con i vincoli di responsabilità di procedimento e relativi tempi di lavorazione delle pratiche.

Per concretizzare, si consideri questo esempio. Arpa Emilia-Romagna deve gestire il catasto della domanda e dell'offerta di energia sul territorio regionale. Con le competenze dettate dalla nuova legge di riordino istituzionale (Lr n. 13/2015), gli impianti di produzione di energia sono sottoposti a pareri autorizzativi Arpa, per cui rientrano nei sistemi informatizzati di gestione delle pratiche. Ma con tempi e modi legati, appunto, al controllo e/o al rilascio di pareri. I gestori del processo sono diversi servizi Arpa distribuiti sul territorio (Servizi territoriali, Sistemi ambientali e il Centro tematico regionale Energia e valutazioni ambientali complesse, Ctr Eva), mentre il detentore del catasto è il Ctr Eva. Inoltre la geolocalizzazione e la creazione di tematismi cartografici sono gestite dal servizio Sistema informativo territoriale. Il modello organizzativo, che resta piuttosto complicato, deve essere proiettato sul fatto che il catasto sia il più possibile completo e aggiornato, ma anche aggiornabile seguendo protocolli efficienti e non sequenziali. Si è scelto di utilizzare uno strumento semplice da utilizzare e di largo uso: il foglio elettronico in modalità *cloud*, che prevede privilegi di accesso differenziati e rigidamente controllati.

Le caratteristiche di partenza più interessanti sono quelle che riguardano i paradigmi della sicurezza informatica. Il foglio elettronico in *cloud*, assegnando opportunamente i privilegi di condivisione, risponde ai requisiti della sicurezza informatica, in termini di:

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

Si noti che nessuno di questi tre requisiti viene soddisfatto da un normale foglio

FIG. 4
MAPPA IMPIANTI

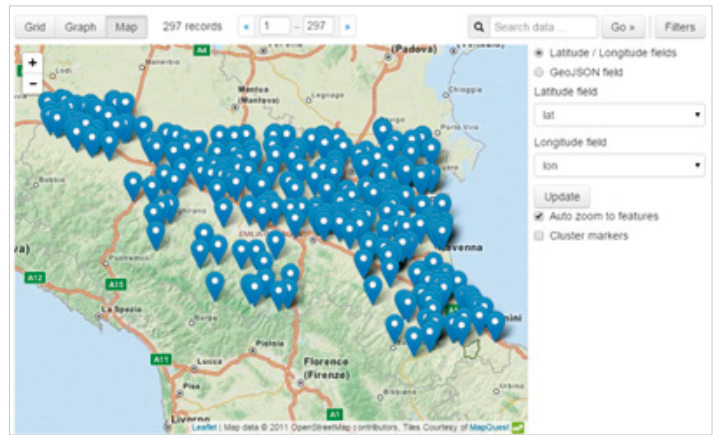


FIG. 5
SCHEDE IMPIANTI

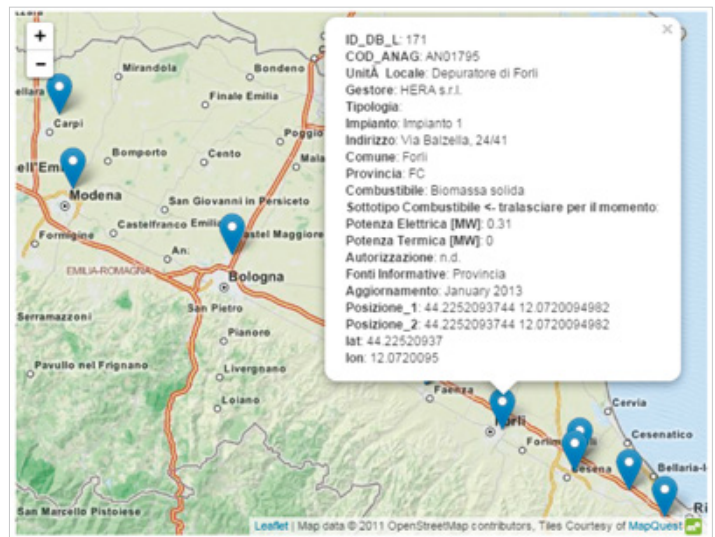


FOTO: MARTINA NOTTE - CC BY-SA 3.0 - WIKIMEDIA COMMONS

elettronico ospitato su un Pc o in una cartella di rete. E nemmeno dai rigidi sistemi gestionali, soprattutto per quanto riguarda la "disponibilità". Viceversa, ciò sarebbe possibile da un sistema centrale, capace di rendere disponibile lo stesso documento a diversi operatori anche dislocati su un vasto territorio, che abbia un sistema di tracciamento automatico delle modifiche (senza accesso anonimo) e con diversi profili, caratterizzati da diversi gradi di accesso ai dati. Alla luce di questo, risulta dunque vantaggioso modificare il consueto flusso di creazione-validazione/trattamento

-diffusione dei dati, allo scopo di consentire una revisione continua del database e dei relativi metadati (in particolare riguardo a un codice di revisione e/o registrazione della data di aggiornamento). Tale nuovo flusso deve tenere conto del processo di aggiornamento dovuto a eventuali ritorni informativi degli utenti (*feedback*). Quindi, da lineare (figura 1) il processo diventa circolare (figura 2) nelle fasi di validazione/trattamento e diffusione attraverso il feedback degli *stakeholder*. Anzitutto si deve definire un punto di accesso al *dataset master* che risponda ai requisiti di sicurezza informatica: questo

è possibile in applicativi mirati residenti su server accessibili in rete e con adeguati profili di accesso.

Potenzialmente molti sistemi possiedono questi requisiti: server cartografici, applicativi gestionali, Dbms. Ma ciò che generalmente manca è un sistema che, oltre ai dati, consenta una completa gestione dei metadati, ovvero delle informazioni di supporto di ogni *dataset*, che ne devono seguire fedelmente le vicende.

Il sistema software Ckan (v. Cattani-Fustini, in *Ecoscienza* 2/2015) permette di gestire tutti gli aspetti di metadatozione dei *dataset*, aggiungendo la possibilità di gestire anche i *dataset* stessi, sia come file, sia come tabelle Db ospitate in un proprio datastore. Dati e metadati possono essere pubblici o privati, consentendo anche una gradualità nella pubblicazione che prevede approssimazioni successive.

Anche i sistemi *cloud* permettono di rispettare i requisiti di sicurezza informatica. In modo particolare, strumenti come Google Spreadsheet lasciano agli operatori la possibilità di lavorare su fogli elettronici già consolidati, ma aggiungendo una serie di controlli di accesso che ne risolvono le criticità. E, soprattutto, attraverso collegamenti "caldi" unidirezionali è possibile tenere costantemente allineato il *dataset* pubblicato con quello in produzione.

Il catasto di impianti aventi "certe caratteristiche" viene definito sulla base dei requisiti informativi richiesti, ad esempio:

- informazioni anagrafiche
- geolocalizzazione
- dati tecnici
- fonte dei dati
- data di aggiornamento
- nome del compilatore.

Ma prima ancora di definire questi dati, può essere utile definire il *dataset* in termini di metadati. Perciò, su Ckan si inseriscono tutte le informazioni che caratterizzano i contenuti del *dataset*, lasciando a una fase successiva la definizione puntuale dei formati e delle modalità di pubblicazione.

Attraverso Google Spreadsheet, se il *dataset* richiede contributi di diversi servizi/aree, è possibile integrare tali contributi in maniera organica e semplice, senza lo scambio di copie di file per posta elettronica e mantenendo traccia delle modifiche effettuate da ogni singolo collaboratore, anche attraverso un attento uso dei privilegi da parte del *process owner* (puntualmente identificato nella scheda metadati di Ckan). Un esempio del sistema descritto, riferito al catasto degli impianti a biomasse, è visibile al link <http://dati.arpa.emr.it/dataset/impianti-energetici-a-biomasse>.

Stefano Cattani, Mauro Rossi

Arpa Emilia-Romagna

GLOSSARIO

Catasto: inventario certificato (statico e aggiornato secondo precise metodologie) di dati descrittivi e quantitativi relativi a un ambito geografico.

Ckan: per dettagli sullo scambio dati, cataloghi open data e piattaforma Ckan si veda la documentazione presente sul sito AGID all'indirizzo www.agid.gov.it/agenda-digitale/open-data/dati-pubblici-condivisione.

Dataset ambientale: insieme di dati coerenti per ambito ambientale o matrice; i dati vengono resi fruibili attraverso una o più risorse.

Inventario: risultati di un'indagine *top-down* o *bottom-up* su una serie di indicatori, spesso finalizzata alla realizzazione di un catasto.

Open data: dati liberamente accessibili a tutti, privi di brevetti o altre forme di controllo che ne limitino la riproduzione.

Risorsa: tabella di dati, singolo file o copertura cartografica relativa a uno specifico *dataset*.

QUALITÀ DELL'ARIA IN EMILIA-ROMAGNA, I DATI 2014 E LE ANTICIPAZIONI 2015

Nel 2014 e nel 2013 è scesa a meno del 20% la popolazione residente in aree dell'Emilia-Romagna esposta al superamento del valore limite fissato dall'Unione europea per la protezione della salute umana relativamente al PM_{10} ($50 \mu\text{g}/\text{m}^3$ di concentrazione media giornaliera da non superare più di 35 volte in un anno); questi valori sono in linea con i valori medi europei riportati dal rapporto dell'Agenzia europea per l'ambiente (Eea) per la popolazione residente nelle aree urbane, anno 2011. Questa percentuale è decisamente inferiore a quella registrata nel periodo 2009-2012, quando la percentuale di popolazione potenzialmente esposta è risultata compresa tra il 50 e il 70%.

Questa situazione è legata all'andamento della concentrazione al suolo di PM_{10} : nel 2014 la concentrazione media annuale di polveri sottili (PM_{10}) ha infatti confermato i minimi storici raggiunti nel 2013; rispettato nel 77% delle stazioni di monitoraggio il limite dei 35 superamenti giornalieri (contro il 28% del 2012 e il 38% del 2013). Rispettati in tutta la regione anche i valori limite annuali per la protezione della salute umana, sia per le polveri sottili sia per quelle ultrasottili. Confermata l'assenza di criticità per gli inquinanti storici quali monossido di carbonio e biossido di zolfo. Bene anche il benzene, la cui concentrazione nell'aria si è progressivamente ridotta e ora è stabilizzata su valori sensibilmente inferiori al limite in tutte le stazioni. In calo anche la media annuale del biossido d'azoto: il limite della media annua è stato rispettato in 36 delle 40 stazioni della rete regionale di monitoraggio. Resta elevata e compresa tra il 70 e 90% la percentuale di popolazione residente in aree della regione, dove la concentrazione di ozono supera il valore obiettivo ($120 \mu\text{g}/\text{m}^3$,



massima delle medie mobili su 8 ore da non superare più di 25 volte anno). La concentrazione di ozono risulta tuttavia in leggera diminuzione dal 2011, anche se continua a registrare valori superiori ai limiti.

I dati della qualità dell'aria in tempo reale

Arpa Emilia-Romagna pubblica quotidianamente i dati relativi alla qualità dell'aria sul proprio sito web (www.arpa.emr.it). In particolare sono disponibili:

- PM_{10} , *statistiche riepilogative* (superamento della soglia di $50 \mu\text{g}/\text{m}^3$ per più di 35 giorni l'anno)
- *situazione riassuntiva regionale* (dati giornalieri delle misure effettuate nelle stazioni di monitoraggio della rete regionale ed elaborazioni statistiche).

Nei primi 11 mesi (gennaio-novembre), l'inquinante che ha maggiormente destato preoccupazioni è stato l'ozono, migliore la situazione per il PM_{10} , seppure ancora critica, con superamento del limite massimo di 35 giorni anno oltre i $50 \mu\text{g}/\text{m}^3$ in 9 stazioni su 43. Buona la situazione per il biossido d'azoto, con tutte le stazioni monitorate entro i valori limite.