

## Convergenza OpenData/INSPIRE, data mining e usabilità nel nuovo sistema di pubblicazione dei dati ambientali ARPAT

Cinzia Licciardello (\*), Camillo Berti (\*\*), Marco Bazzani (\*\*\*), Stefano Mignani (\*\*\*\*)

(\*) ARPAT, via Porpora 21, +39 055/320.6413, +39 055/320.6410, c.licciardello@arpat.toscana.it

(\*\*) ARPAT, via Porpora 21, +39 055/320.5251, +39 055/320.6410, c.berci@arpat.toscana.it

(\*\*\*) ARPAT, via Porpora 21, +39 055/320.6466, +39 055/320.6466, m.bazzani@arpat.toscana.it

(\*\*\*\*) ARPAT, via Porpora 21, +39 055/320.5251, +39 055/320.6359, s.mignani@arpat.toscana.it

### Introduzione

*Garantire una reale fruibilità dei dati ambientali da parte delle pubbliche amministrazioni e dei cittadini sia secondo le specifiche nazionali (Open Data) che comunitarie (INSPIRE) costituisce spesso una 'sfida nella sfida', a causa di caratteristiche peculiari quali numerosità e multidimensionalità. Presso il Sistema Informativo Regionale dell'Ambiente della Toscana (SIRA) è stato realizzato un sistema integrato di catalogazione e distribuzione dei dati ambientali disponibili presso ARPAT in conformità a quanto previsto dalle normative Open Data (nazionale) ed INSPIRE (comunitaria), con il duplice obiettivo di garantire agli utenti la maggiore usabilità possibile e di semplificarne al massimo le attività di gestione. Un archivio centralizzato (catalogo dei dati disponibili) viene utilizzato sia per la generazione automatica della metainformazione secondo lo standard ISO19139 che per la distribuzione dei dati tramite servizi INSPIRE (WMS, WMTS, WFS, CSW) e un sistema di API pubbliche, mentre l'usabilità del sistema lato utente è garantita da un'interfaccia web per la ricerca, la navigazione, il data mining e lo scarico dei dati.*

*Making environmental data easily available to public citizens and public companies by mean of both national (Open Data) and EU (INSPIRE) specifications can be a challenging task, due to the large amount of records and dimensions involved. The Environmental Regional Information System of Tuscany (SIRA) has built an easy-to-use centralized system based on a unique metadata archive: such an archive is used for both metadata automatic generation and access services configuration by mean of public APIs and INSPIRE services (WMS, WMTS, WFS, CSW). For better usability a UI web interface for data search, navigation, data mining and download has been developed.*

### Requisiti normativi e standard tecnici per la pubblicazione dei dati ambientali

La recente normativa nazionale sull'apertura dei dati in possesso delle pubbliche amministrazioni (normativa Open Data e sulla trasparenza delle pubbliche amministrazioni) è venuta ad affiancarsi alla preesistente normativa comunitaria in tema di banche dati di interesse ambientale (direttiva 2007/2/EC [INSPIRE]), che definisce ambiti e modalità di pubblicazione per i dati territoriali di interesse ambientale. La specificità dei dati territoriali, la cui gestione richiede un elevato livello di specializzazione, ha sinora comportato l'adozione di soluzioni tecnologiche e architetture per la loro pubblicazione (infrastrutture di dati geografici o geoportali) che si solito differiscono da quelle generalmente adottate per la pubblicazione di dati aperti privi di informazione geografica associata (portali open data con supporto alla ricerca semantica).

L'utilizzo di strumenti diversi per il raggiungimento di finalità analoghe genera inevitabilmente un surplus di sistemi informatici la cui gestione si complica notevolmente al crescere della numerosità delle banche dati pubblicate, complicando anche la creazione di un catalogo unico di tutte le banche dati, geografiche e non, gestite da una pubblica amministrazione: risulta di gran lunga preferibile un

approccio unitario al problema della distribuzione dei dati gestiti da un'organizzazione, che presentino o meno una componente geografica.

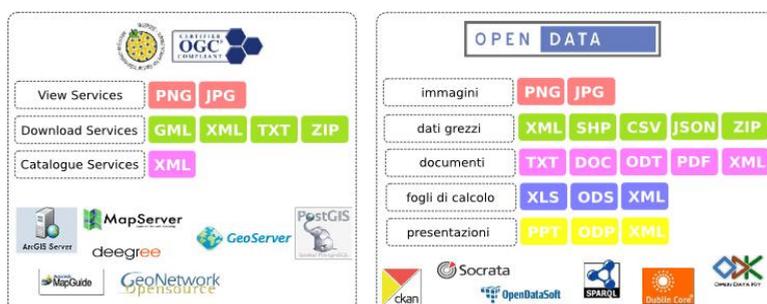


Figura 1 □ Formati di dati e soluzioni tecnologiche di utilizzo comune conformi alle specifiche INSPIRE e Open Data (<http://opendatahandbook.org/it/appendices/file-formats.html>).

### INSPIRE e OpenData: una convergenza possibile

La direttiva INSPIRE prevede la distribuzione di dati geografici tramite varie tipologie di servizi web aderenti a standard OGC:

- *View Services* per la rappresentazione grafica: *Web Map Service* (WMS) e *Web Map Tile Service* (WMTS)
- *Download Services* per lo scarico dei dati: *Web feature Services* (WFS) e *Sensor Observation Service* (SOS)
- *Discovery Services* per la ricerca dei dati disponibili: *Web Catalogue Service* (CSW)

Si può facilmente notare come tutti i dati disponibili tramite servizi pubblici WFS (dati geografici) e SOS (dati provenienti da reti di monitoraggio associabili a oggetti geografici) e consultabili a partire da servizi di catalogo CSW rispettano i requisiti OpenData livello 4:

- sono disponibili in formato aperto, strutturato e standard W3C (gml o xml)
- sono raggiungibili raggiungibili a specifiche URL e liberamente scaricabili via web
- e' disponibile un catalogo dei metadati
- e' disponibile un'interfaccia applicativa (API) di accesso con possibilita' di filtraggio dei dati

Di conseguenza il sistema di pubblicazione di dati territoriali secondo la direttiva INSPIRE, ove disponibile (sviluppato internamente oppure disponibile come software Open Source), puo' essere utilizzato in larga parte per esporre anche dai di tipo non geografico in formato xml con minime modifiche al codice sorgente. L'utilizzo del formato XML, vera e propria "lingua franca" per l'interscambio dei dati, garantisce di per sè il rispetto della filosofia alla base della normativa Open Data, e costituisce il primo passo verso l'interoperabilità semantica tra dati.



Figura 2 □ Formati comuni alla normativa nazionale (Open Data) e comunitaria (INSPIRE).

### Metadati e servizi INSPIRE: necessità di una maggiore integrazione

Anche in ambito INSPIRE per motivi organizzativi e gestionali i due archivi dei metadati descrittivi dei dati territoriali e dei metadati descrittivi dei servizi web (metadati descrittivi dei servizi e dei contenuti) vengono gestiti separatamente, anche a causa della scarsa disponibilità di soluzioni integrate: anche in questo caso una più stretta integrazione tra gli archivi semplificherebbe non poco l'intero workflow di pubblicazione, consentendone un maggiore livello di automazione. Un archivio unico dei metadati degli archivi e dei servizi potrebbe essere utilizzato come motore dell'intero sistema di pubblicazione (servizi di catalogo e servizi di distribuzione dei dati), collegando informazioni descrittive dei dati e informazioni necessarie al loro recupero, andando a costituire il nucleo centrale di un'architettura centrata sui metadati.

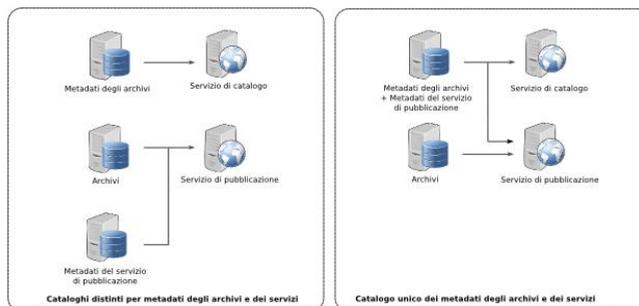


Figura 3 □ Architetture a confronto per la metainformazione relativa a dati territoriali e servizi.

Architettura centrata sui metadati, aderenza agli standard in essere e riuso

L'esperienza maturata dal settore nel corso di un decennio di utilizzo di diversificati strumenti di pubblicazione di una quantità crescente di dati ha evidenziato una serie di elementi sistemistici, architetture e applicativi necessari per garantire la massima semplicità d'uso sia da parte dei gestori che da parte degli utenti di un sistema ideale di pubblicazione di dati territoriali e non:

- unicità
- scalabilità
- compatibilità con le specifiche INSPIRE e Open Data
- disponibilità di un catalogo centralizzato e di API pubbliche
- usabilità dell'interfaccia
- possibilità di effettuare operazioni di data mining lato utente

Per salvaguardare gli investimenti in essere, in un'ottica di riuso dei sistemi ITC standard in ARPAT e di piena valorizzazione delle professionalità del *team* di sviluppatori, sistemisti e gestori delle banche dati, si è deciso di non affiancare soluzioni non compatibili o difficilmente integrabili, per quanto 'chiavi in mano', optando per la realizzazione di soluzioni integrate con l'infrastruttura ARPAT. Per questo motivo è stato sviluppato in proprio da parte del Sistema Informativo Regionale dell'Ambiente della Toscana un sistema CMS (*Content Management System*), sfruttando al meglio le varie tecnologie già disponibili, in modo da poter disporre di un'unico sistema di pubblicazione sia Open Data che INSPIRE compatibile.

L'architettura sviluppata presso il settore SIRA consente, a partire dalla definizione di informazioni sintetiche relative ai dataset da pubblicare (tabella parametri dataset) e dalla definizione di un set minimo di metainformazione, di pubblicare ciascun dataset sia come servizio INSPIRE che come dato scaricabile in formato aperto tramite accesso diretto a un *Uniform Resource Locator* (URL) o tramite API pubblica, garantendo al contempo la consultazione e lo scarico della metainformazione.

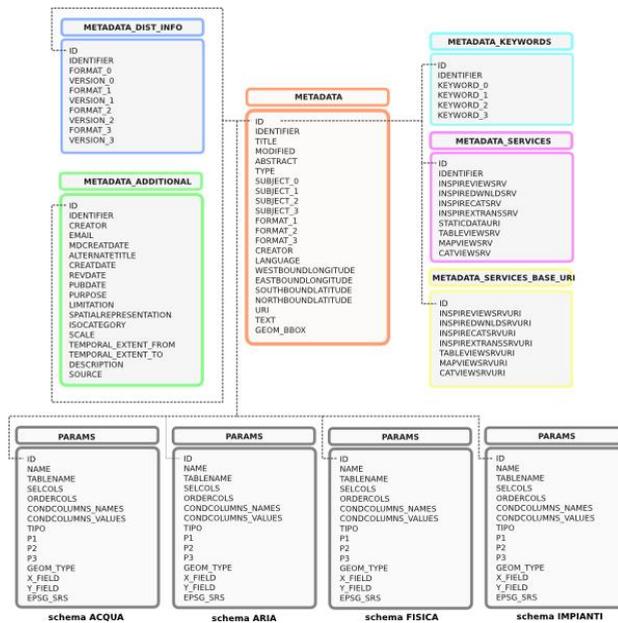


Figura 4 □ Schema del sistema CMS per la pubblicazione dei dati.

### Tecnologie e standard utilizzati

Il sistema RDBMS Oracle, pur in assenza di licenza *spatial*, si è dimostrato un ottimo strumento per l'archiviazione e il recupero di dati geografici di varie tipologie (coppie di coordinate, linee, poligoni) grazie alle funzionalità rese disponibili dal componente di base *locator*, ed è stato utilizzato per l'esportazione dei dati in formato XML tramite procedure PL/SQL dinamiche. Il sistema *APEX Application Express*, già in uso presso il settore SIRA per lo sviluppo di semplici applicativi e report a partire da banche dati Oracle, è stato utilizzato come motore di servizi di accesso REST configurabili tramite query o processi PL/SQL a partire da dati archiviati su Oracle: tali tecnologie risultano sufficientemente flessibili per realizzare un'unico sistema di pubblicazione sia Open Data che INSPIRE compatibile, e sono disponibili gratuitamente sia nel caso di licensing base che per la versione gratuita OracleXE.

I parametri per le procedure dinamiche vengono recuperati all'interno di un applicativo PHP dalle chiamate ai servizi INSPIRE (nome del *dataset* ed eventuali filtri geografici o alfanumerici), oltre che dai parametri di configurazione del *dataset* e dai metadati esposti come servizi web tramite APEX (mod-plsql) e successivamente passati all'applicazione APEX. Il dato XML viene recuperato all'interno dell'applicazione APEX e restituito all'applicativo PHP, che provvede a trasformarlo nel formato richiesto (GML per servizi WFS, SVG per il servizio WMS, CSV/XML/KML/JSON/GeoJSON per l'accesso diretto secondo le specifiche Open Data).

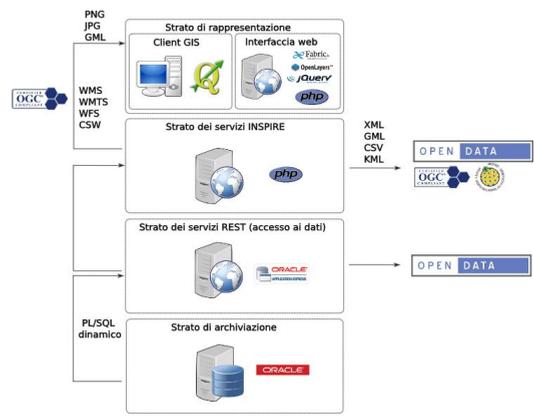


Figura 5 □ Infrastruttura tecnologica del sistema di pubblicazione dei dati.

Particolare attenzione e' stata posta agli aspetti di usabilita', sia nei confronti degli utenti abituali dei dati territoriali (specialisti nella gestione dei dati e tecnici ARPAT competenti), tradizionalmente piu' interessati all'utilizzo in ambiente GIS, che nei confronti degli utenti incaricati delle attività di reporting ambientale, piu' interessati all'utilizzo di semplici interfacce web per il data mining e lo scarico di subset di grosse moli di dati.

Per la prima tipologia di utenti e' stato testato il funzionamento dei servizi INSPIRE con il software QGIS versione 2.4, che costituisce la soluzione standard presso ARPAT per l'accesso ai dati geografici, mentre per la seconda tipologia di utenti e' stata realizzata una pagina di ricerca che consente l'individuazione dei dataset disponibili tramite il servizio di catalogo dei metadati e l'estrazione dei subset di interesse in base a selezioni per area geografica (provincia, comune, altri oggetti geograficamente definiti), per indicatore (parametro monitorato/indicatore normato, tipologia di impianto...) e/o per periodo (anno).

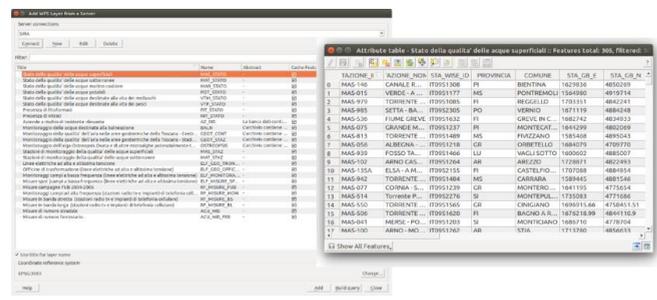


Figura 6 □ Accesso ai dati territoriali tramite QGIS.

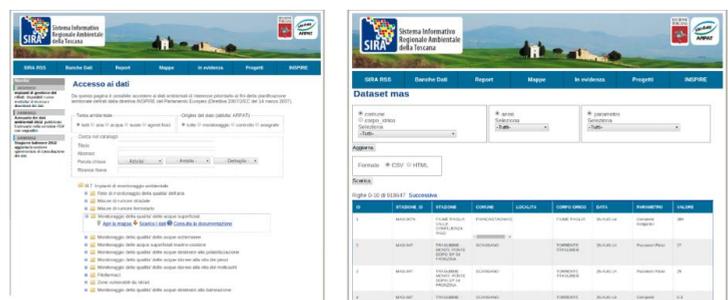


Figura 7 □ Accesso ai dati tramite interfaccia web.

## Conclusioni e prospettive future

Il sistema di pubblicazione di dati ambientali sviluppato presso il settore SIRA ha consentito di evitare l'adozione di nuove tecnologie in aggiunta a quelle già esistenti, con un notevole risparmio di risorse umane e strumentali, focalizzando l'attività sul 'core business' delle attività di apertura dei dati verso l'esterno con una maggiore attenzione all'organizzazione e alla catalogazione delle informazioni disponibili. L'architettura adottata risulta facilmente estendibile e manutenibile, garantendo inoltre un ampio riutilizzo di oggetti e l'ottimizzazione delle operazioni di configurazione e documentazione del sistema.

Gli utenti del portale dei dati ambientali gestiti dall'Agenzia Regionale per la Protezione Ambientale della Toscana possono adesso disporre di un'interfaccia di semplice utilizzo per la navigazione e lo scarico dei dati ambientali, basata su un sistema che oltre a preservare gli investimenti in atto consente una piena valorizzazione delle tecnologie e delle risorse umane, garantendo al contempo un notevole risparmio in termini di TCO (Total Cost of Ownership).

Molto lavoro rimane ancora da fare nell'ottica di raggiungere il livello 5 di qualità dei dataset esposti per poter garantire una reale interoperabilità semantica tra i dati ARPAT e altri dati pubblici a questi collegabili. In quest'ottica, in futuro si prevede di concentrare le attività future sui seguenti aspetti:

- Armonizzazione del contenuto dei dati con le ontologie definite a livello comunitario e nazionale
- Arricchimento dei dati con riferimenti incrociati ad altri oggetti esterni disponibili sia presso ARPAT che presso altre pubbliche amministrazioni (LOD: *Linked Open Data*)
- Realizzazione di un *endpoint* SPARQL per l'interrogazione semantica dei dati

## Bibliografia

Ciasullo G., Lodi G., Maccioni A., Rotundo A. (2014), "Linee Guida Nazionali Per La Valorizzazione Del Patrimonio Informativo Pubblico (anno 2014)"  
[http://www.agid.gov.it/sites/default/files/linee\\_guida/patrimoniopubblicolg2014\\_v0.7finale.pdf](http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.7finale.pdf)

Licciardello C. (2013), "Environmental SDI in Tuscany: State of The Art and Future Developments", INSPIRE conference 2013, Firenze  
[http://inspire.ec.europa.eu/pr\\_searchx.cfm?i=5&id\\_search=10670](http://inspire.ec.europa.eu/pr_searchx.cfm?i=5&id_search=10670)

Hillenius J. (2013), "Tuscany's environmental organisation prefers open source", Open source observatory News  
<https://joinup.ec.europa.eu/community/osor/news/tuscanys-environmental-organisation-prefers-open-source>

Cornero A. (a cura di), Belisario E., Epifani S. (2011), "Vademecum Open Data"  
<http://www.dati.gov.it/sites/default/files/VademecumOpenData.pdf>

Růžička J. (2011), "Catalogue Services Lite - INSPIRE Profile", Symposium GIS Ostrava  
[http://gis.vsb.cz/GIS\\_Ostrava/GIS\\_Ova\\_2011/sbornik/papers/Ruzicka.pdf](http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2011/sbornik/papers/Ruzicka.pdf)

Berti C. Licciardello C (2008), "Verso l'infrastruttura di dati spaziali: progetti e servizi del Sistema Informativo Regionale Ambientale della Toscana", in Atti del VII workshop 'Beni ambientali, culturali e GIS', Trieste

Berti C. Licciardello C (2005), "Verso la costruzione di un'infrastruttura di dati spaziali di interesse ambientale con strumenti open source in Toscana", in Atti della IX conferenza ASITA, Catania  
<http://www.attiasita.it/Asita2005/Pdf/0310.pdf>

## Manualistica

Kothuri R.V., Godfrind A., Beinat E. (2007), *Pro Oracle Spatial for Oracle Database 11g*, Apress, New York, NY

<http://www.apress.com/9781590598993>

Van Zoest M., Van Der Plas M., (2013), *Oracle APEX Cookbook: Second Edition*, Packet Publishing, Birmingham, UK

<https://www.packtpub.com/application-development/oracle-apex-cookbook-second-edition>

Aust D., Giffy D'Souza M., Gault D., Gielis D., Hartman R., Hichwa M., Kennedy S., Kubicek D., Mattamal R., McGhan D., Mignault F., Nielsen A., Scott J. (2011), *Expert Oracle Application Express*, Apress, New York, NY

<http://www.apress.com/9781430235125>

Fox T., Scott Spendolini S., Scott J. (2011), *Pro Oracle Application Express 4, 2nd Edition*, Apress, New York, NY

<http://www.apress.com/9781430234944>

Gupta S. K. (2012), *Oracle Advanced PL/SQL Developer Professional Guide*, Packet Publishing, Birmingham, UK

<https://www.packtpub.com/big-data-and-business-intelligence/oracle-advanced-plsql-developer-professional-guide>

Juneau J., Arena M. (2010), *Oracle PL/SQL recipes*, Apress, New York, NY

<http://www.apress.com/9781430232070>

## Sitografia

Portale del Sistema Informativo Regionale dell'Ambiente della Toscana (SIRA)

<http://sira.arpat.toscana.it>

Riferimenti nazionali Open Data

<http://opendatahandbook.org/it/>

Esempi applicativi

<http://spcdata.digitpa.gov.it/index.html>

<http://www.opendatanetwork.it/>

Progetto INSPIRE: *implementing rules*, riferimenti normativi, eventi

<http://inspire.ec.europa.eu/>

Piattaforma JoinUp per

<https://joinup.ec.europa.eu/>

Geospatial Data and the Semantic Web

<http://geoknow.eu/Welcome.html>

Standard W3C: *semantic web*

<http://www.w3.org/standards/semanticweb/>

Standard W3C: formato XML

<http://www.w3.org/standards/xml/>

Standard W3C: servizi web

<http://www.w3.org/standards/webofservices/>

Standard Open GIS Consortium (OGC): WMS, WFS, CSW, SOS

<http://www.opengeospatial.org/standards>

Librerie per map clients e HTML5 canvas

<http://www.openlayers.org>

<http://fabricjs.com/>

Oracle: componente Locator e APEX

<http://www.oracle.com/it/products/database/options/spatial/index.html>

<http://www.oracle.com/us/products/database/options/spatial/039946.pdf>

<https://apex.oracle.com/i/>