

Formati aperti e metadati per il riutilizzo e la diffusione dei dati pubblici

Il presente allegato illustra brevemente la definizione di formato aperto ed elenca i formati aperti più comuni ed i metadati per la pubblicazione dei dati aperti dall'azienda Ospedaliera della provincia di Lecco.

Formati aperti

Come indicato nel documento delle Linee Guida, i dati di tipo aperto devono essere “accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti”, cioè resi pubblici, documentati esaustivamente e senza vincoli all'implementazione; “sono adatti all'utilizzo automatico da parte di programmi per elaboratori” (machine readable) e sono “provvisi dei relativi metadati”. L'accessibilità da parte di altri programmi di elaborazione automatica è ottenuta attraverso l'uso di formati aperti tali da permettere di massimizzare le loro possibilità di riutilizzo e interoperabilità. Inoltre il processo di creazione e pubblicazione dei dati è molto più semplice quando il formato di dato aperto viene adottato direttamente alla sorgente invece della riconduzione successiva con delle attività di trasformazione e modifica. Per contro si osserva anche che, esistendo grandi quantità di dati in formati standard “de facto”, ovvero formati non aperti ma ugualmente leggibili e visualizzabili con software proprietari e open source (come per esempio il formato Microsoft Excel antecedente il 2003), può risultare necessario pubblicare i dati anche in questi formati.

A titolo di esempio, la tabella 1 fornisce un primo elenco dei formati aperti maggiormente diffusi nell'ambito dell'amministrazione provinciale. La tabella, nella prima colonna, elenca il nome, l'acronimo e una breve descrizione del formato. La seconda colonna contiene il tipo di dato pubblicato dal formato aperto. La terza colonna indica l'estensione del file normalmente utilizzata per questo formato. Da notare che quando un insieme di dati aperti è pubblicato in più file (es. ESRI shapefile) questi possono essere distribuiti all'interno di un file archivio compresso (es. “.zip”). In questo caso è necessario indicare il formato dei dati contenuti nell'archivio.

Tabella 1. Formati aperti adatti alla pubblicazione dei dati aperti

Nome (Acronimo) Descrizione	Tipo di Dato	Estensione del file
Comma Separated Value (CSV) Formato testuale per l'interscambio di tabelle, le cui righe corrispondono a record e i cui valori delle singole colonne sono separati da una virgola (o punto e virgola)	Dato tabellare	.csv
Geographic Markup Language (GML) Formato XML utile allo scambio di dati territoriali di tipo vettoriale	Dato geografico vettoriale	.gml

<p>Keyhole Markup Language (KML)</p> <p>Formato basato su XML creato per gestire dati territoriali in tre dimensioni nei programmi Google Earth, Google Maps</p>	Dato geografico vettoriale	.kml
<p>Open Document Format per dati tabellari (ODS)</p> <p>Formato per l'archiviazione e lo scambio di fogli di calcolo</p>	Dato tabellare	.ods
<p>Resource Description Framework (RDF)</p> <p>Basato su XML, e' lo strumento base proposto da World Wide Web Consortium (W3C) per la codifica, lo scambio e il riutilizzo di metadati strutturati e consente interoperabilità tra applicazioni che si scambiano informazioni sul Web</p>	Dato strutturato	.rdf
<p>ESRI Shapefile (SHP)</p> <p>Lo Shapefile ESRI è un popolare formato vettoriale per sistemi informativi geografici. Il dato geografico viene distribuito normalmente attraverso tre o quattro files (se indicato il sistema di riferimento delle coordinate). Il formato è stato rilasciato da ESRI come formato (quasi) aperto</p>	Dato geografico vettoriale	.shp, .shx, .dbf, .prj
<p>Tab Separated Value (TSV)</p> <p>Formato testuale per l'interscambio di tabelle, le cui righe corrispondono a record e i cui valori delle singole colonne sono separati da un carattere di tabulazione</p>	Dato tabellare	.tsv
<p>Extensible Markup Language (XML)</p> <p>E' un formato di markup, ovvero basato su un meccanismo che consente di definire e controllare il significato degli elementi contenuti in un documento o in un testo attraverso delle etichette (markup)</p>	Dato strutturato	.xml

Metadati

Nello specifico ambito dei dati aperti, i metadati sono intesi come i “dati associati ai dati” tali da informare gli utilizzatori del significato del dato stesso. La definizione stessa di metadato evidenzia l'ampiezza dei possibili metadati associabili ad un insieme di dati (*dataset*), ed è per questo che è opportuno identificare un insieme minimo di informazioni di base che possano essere utilizzate ai fini della ricerca, dell'accesso e dell'utilizzo del dato pubblicato. Diventa altresì essenziale che venga posta la massima cura nel fornire tali informazioni all'utilizzatore del dato. Alcuni dati di settore includono anche il metadato già codificato in formato standard,

ma nella maggior parte dei casi è necessario definire una documentazione di minima non essendo disponibili standard precisi in tale senso. Per tale motivo è qui di seguito definito un set minimo di metadati considerati indispensabili per la pubblicazione dei dati aperti da parte dell’Azienda Ospedaliera della provincia di Lecco.

La tabella 2 definisce il set minimo di metadati richiesto per la pubblicazione dei dataset di dati aperti dell’Azienda Ospedaliera della provincia di Lecco. Ulteriori metadati opzionali, inclusi quelli illustrati in tabella 3, possono essere specificati per il dataset.

Tabella 2. Metadati: set minimo

Metadato	Descrizione
Titolo	Nome del dataset comprensibile da un utente generico. E' il nome
Titolare	Ente titolare del dataset*
Referente	Struttura o Unità organizzativa referente del dataset*
Contatto	Indirizzo email per la richiesta di informazioni
Descrizione	Descrizione testuale del contenuto informativo del dataset
Categorie	Categoria o categorie alle quali il dataset appartiene*
Tag/Parole chiave	Lista di termini associati al dataset che lo descrivono, e rendono possibile la sua classificazione e la sua ricerca
Documentazione tecnica	Indirizzo o indirizzi delle pagine web che contengono informazioni utili alla comprensione del contenuto del dataset
Descrizione campi	<p>Se non specificato nella descrizione del dataset o nella “Documentazione tecnica”, è necessario descrivere in questo metadato i campi (attributi) delle informazioni contenute nel dataset. Nel caso di dato tabellare, gli attributi corrispondono alle colonne della tabella. Le informazioni da specificare includono:</p> <ul style="list-style-type: none"> ● Nome completo. E' il nome che viene mostrato all'utente quando consulta il dataset (es. “Descrizione”) ● Nome breve. E' il nome dell’attributo utilizzato nel dataset, normalmente un’abbreviazione del nome completo (es. “Desc”) ● Descrizione. Illustra il significato del campo, eventuali relazioni con altri dataset, come il campo è stato generato, etc. (es. “Codice ISTAT del comune amministrativo”) ● Tipologia dell'attributo (testo, numero, data, etc.)* ● Formato dell'attributo (es. unità di misura o struttura della data)

Copertura <u>Temporale</u> (Data di inizio)	Data di inizio validità
Copertura Temporale	Data di fine validità
Aggiornamento	Frequenza di aggiornamento (es. "mensile", "annuale", "giornaliera")*
Data di pubblicazione	Data in cui il dataset è reso disponibile on-line
Data di aggiornamento	Data di ultima modifica del dataset
Formato	Formato di pubblicazione (standard) del dataset*
Codifica Caratteri	Codifica dei caratteri utilizzata (es. "latin-1", "PC-850")*
Dimensione	Dimensione in byte (o multipli di byte) del dataset
URI permanente	Identificatore univoco del dataset

* Lista controllata

Tabella 4. Metadati opzionali

Metadato	Descrizione
Autore	Struttura o persona che ha creato il dataset
Email autore	Email dell'autore
Data di creazione	Data di creazione del dataset
URL sito	Link web alla pagina o al portale istituzionale del dataset
Nome	Nome del file che memorizza il dataset (es. "ammcom.shp")