

OPEN PUBLIC LIBRARY

KIT di riuso Fase B – Trasferimento e adozione della buona pratica

B3. Strumenti tecnologici

B3.2 La soluzione tecnologica

B3.2 Modalità di adozione

Data rilascio: 10/12/2018

Versione: 1.0

Sommario

Premessa	2
1. La soluzione tecnologica.....	2
2. Modalità di adozione	4

Premessa

Il presente documento descrive gli **strumenti tecnologici del kit di riuso per la fase B** – Trasferimento e adozione della buona pratica. Gli strumenti tecnologici, come previsto dall'Avviso OCPA2020¹, consistono in: codice software sorgente compilato della soluzione, con la debita documentazione a corredo, ivi inclusa la descrizione dell'infrastruttura tecnologica della soluzione; documento di licenza d'uso individuata per il software che ne titola la fruibilità come riuso di software pubblico.

NB: tutto quanto inserito negli strumenti tecnologici deve risultare coerente con il CAD e con le Linee Guida sull'acquisizione e sul riuso di software per le pubbliche amministrazioni definite da AgID. In particolare, occorre fare riferimento alla Guida alla pubblicazione di software come open source di AgID (Allegato A alle linee guida) e alla Guida alla manutenzione di software open source (Allegato B alle linee guida).

1. La soluzione tecnologica

La realizzazione del sistema di *Data analysis* si compone di alcuni moduli indipendenti, sviluppati separatamente e reperibili su <https://github.com/Provincia-di-Brescia>.

Si tratta di semplici script php e python, che non richiedono particolari condizioni riguardo la configurazione del sistema e che possono essere facilmente adattati a contesti differenti rispetto a quello finora sperimentato.

Il flusso funzionale prevede:

- La selezione, raccolta e raggruppamento dei dati presenti nel database di produzione del sistema informativo
- La loro scrittura in un db dedicato all'attività di *data analysis* (OLAP). Ciò assicura la storicizzazione dei dati (ossia la preservazione della loro coerenza nel corso del tempo) e la loro rapida reperibilità
- L'olap è accessibile in tre modalità
 - Tramite web-Api
 - Predisponendo un sistema di *data visualization* che acceda direttamente all'olap attraverso le api del DBMS
 - Attivando un *daemon* che trasferisca, in modalità *push*, i dati verso il sistema di *data visualization* che risiede nel cloud.

Nel dettaglio:

1. la soluzione è utilizzabile solo per LMS (Library Management Systems) ClavisNG (<https://github.com/comperio/clavisng-sources>).
2. lo script OLAP.php può essere messo in funzione su qualunque host (OLAP-host) in grado di accedere, in sola lettura, al db di produzione di Clavis. Attualmente OLAP-host è un server linux ma non dovrebbero presentarsi difficoltà significative impiegando sistemi operativi differenti
3. OLAP.php riceve la propria configurazione dal file `olapSettings.json`, documentato nel repository GitHub

¹ OCPA – OpenCommunityPA2020: primo Avviso pubblico per interventi volti al trasferimento, evoluzione e diffusione di buone prassi fra Pubbliche Amministrazioni.

4. Da OLAP-host, lo script deve poter accedere, in scrittura, a un DBMS Mysql o MariaDB, sul quale verrà generato l'OLAP. Non sussistono vincoli circa il sistema operativo che ospiterà il DBMS; attualmente si utilizza Ubuntu-Linux
5. Gli script php per la gestione delle web-API sono stati progettati per il web-server *apache*. Per garantire il loro funzionamento è necessario che possano accedere, in sola lettura, al DBMS sul quale è generato l'OLAP
6. Se si desidera, così come è accaduto fino a questo momento, utilizzare un sistema di *Data Visualization* presente sul cloud e facilmente configurabile (nello specifico, Google DataStudio), è necessario porre in esecuzione quotidiana lo script Python che, leggendo i dati dal DBMS sul quale è ospitato l'OLAP, ne esegua il "push" verso lo spazio di archiviazione utilizzato per la costruzione dei web-reports.

Riguardo la struttura dei database implicati nella soluzione:

1. Il db Clavis deve essere ben padroneggiato ai fini della corretta configurazione del sistema. È opportuno che chi provvederà alla messa a punto dell'olap abbia accesso, almeno in lettura, al db di produzione cui fa riferimento il sistema informativo della rete bibliotecaria. Addirittura, sarebbe consigliabile che l'addetto alla configurazione dell'OLAP sia autorizzato anche alla scrittura delle tabelle del db; infatti, si dà frequentemente il caso in cui, nel corso dell'attività di predisposizione delle query che alimenteranno l'OLAP, emerga la necessità di operare interventi di bonifica e miglioramento della qualità dei dati. Lo schema del db Clavis è facilmente interpretabile e, ai fini della scrittura delle query sql più frequentemente richiesta dall'attività di *data analysis*, si dà indicazione del significato delle tabelle più significative:
 - a. *Manifestation* contiene i dati relativi alle notizie bibliografiche, ossia la descrizione standardizzata delle caratteristiche semantiche e semiotiche dei documenti presenti nelle biblioteche
 - b. *Item* contiene le informazioni relative agli esemplari presenti nelle biblioteche. Il campo *manifestation_id* fornisce il legame alla manifestation corrispondente (naturalmente si tratta di un legame uno a molti: una manifestation può essere legata a più item)
 - c. *Authority* contiene gli elementi di accesso che la ricerca bibliografica impiega per raggiungere le notizie bibliografiche: autori, editori, luoghi, parole chiave, ... Authority intrattiene una relazione molti a molti con sé stessa (tramite *I_authority*) e una relazione molti a molti con manifestation (tramite *I_authority_manifestation*)
 - d. *Patron* contiene i dati relativi agli utenti del servizio bibliotecario
 - e. *Loan* registra i prestiti eseguiti dagli utenti. Il campo *patron_id* pone in relazione il prestito con l'utente che l'ha eseguito
 - f. *Library* contiene le informazioni relative alle biblioteche che compongono il sistema bibliotecario. Sia la tabella item, sia la patron, sia la loan, esprimono la loro relazione con le biblioteche attraverso il campo *library_id*.
2. Il database automaticamente generato dallo script OLAP.php ha uno schema che dipende dal modo in cui le misure sono calcolate. Infatti, lo script elabora il file di configurazione *olapSettings.json* che elenca, oltre ad altre informazioni, le query sql utilizzate per estrarre i dati dal db di produzione. I campi previsti dall'output di queste query vengono automaticamente utilizzati per la creazione delle tabelle dell'olap, ove ognuno dei "fatti" sui quali si eseguono le misure è rappresentato da una

specificata

tabella.

OLAP.php aggiunge a quelli prodotti dalla query alcuni campi standard, dedicati alla data a cui la misura si riferisce.

2. Modalità di adozione

L'adozione della soluzione è, dal punto di vista tecnico, assai semplice e ha un solo vero prerequisito: la possibilità di accedere, almeno in lettura, al db di produzione del sistema ClavisNG con il quale il sistema bibliotecario è gestito.

La configurazione dei server sui quali eseguire gli script che compongono la soluzione può essere adattata facilmente alle disponibilità del riusante. Si suggeriscono due possibili approcci:

1. Installazione "in house". Si consiglia di alloggiare il DBMS che ospiterà l'OLAP nella MZ. Il server incaricato della gestione del DBMS potrebbe essere utilizzato sia per l'esecuzione dello script OLAP.php sia per l'esecuzione dello script python che si occuperà (nel caso si scelga questa soluzione) di trasportare i dati in modalità "push" verso un sistema di *Data visualization* esterno al dominio di riferimento. Un host della DMZ dovrebbe essere destinato all'esecuzione delle API
2. Uso della soluzione in "cloud". Se il sistema bibliotecario che desidera utilizzare la soluzione non ritenesse opportuno adottare la soluzione "in house" (per scarsità di risorse tecniche o per il desiderio di sperimentare la soluzione prima di procedere alla sua acquisizione sui propri server), potrebbe sottoscrivere un accordo con una delle istituzioni che già la utilizzano affinché tutte le configurazioni vengano gestite da operatori già esperti. In questo caso va studiata la migliore modalità di accesso ai dati del sistema informativo di produzione: a valle degli adempimenti previsti dal GDPR le soluzioni possibili potrebbero essere o l'accesso al database tramite vpn o ipsec oppure il trasferimento periodico di una copia del database da analizzare.