Creazione di un Archivio

La creazione di un nuovo archivio eXtraWay può avvenire contestualmente alla realizzazione del suo modello dati tramite la configurazione di una DTD ovvero in forma libera. Per quanto il primo sistema è quello che risulta più corretto dal punto di vista formale, la pratica ci dice che gli archivi vengono normalmente creati in forma libera, specialmente quando il progetto che li riguarda è in divenire. La presente documentazione si sofferma quindi sulla forma libera della creazione degli archivi.

Cos'è un Archivio eXtraWay

Prima di procedere con la spiegazione di come si costituisca un archivio, soffermiamoci ad identificare cosa si intenda per Archivio eXtraWay. Anche al lettore meno attento sarà evidente che i *dati* in un archivio eXtraWay, sono rappresentati da files XML[1)] che contengono in forma ordinata una o più *Unità Informative*.
Essi, in quanto tali, non possono rappresentare da soli un archivio. Perché ciò possa avvenire essi devono essere *Mappati* ovvero interpretati nel loro contenuto e rappresentati appunto attraverso una mappa in grado di identificare entro ciascun file XML ciascuna *Unità Informativa*. In seguito, in corrispondenza ad inserimenti, modifiche e cancellazioni, le *Unità Informative* verranno identificate nella mappa che subirà a sua volta inserimenti, modifiche e cancellazioni. Una volta costituito il Data Base con i dati identificati per mezzo della mappa, sui contenuti delle *Unità Informative* verranno valutati gli indici e costruiti i files a ciò preposti.
Scopo di quest'architettura è quello di concentrare sugli XML il ruolo di vera importanza: la mappa e gli indici possono essere generati/rigenerati in qualsiasi momento.
Per avere una rapida visione d'insieme diremo che:

Tipo di file	Ruolo
.stat.xml[2)] **.lck**	File di *Status* che coordina i contenuti ed il legame tra tutti gli altri files che compongo un archivio e file che consente la regolazione degli accessi
.udp **.udd**	File di mappa. Contengono le informazioni che consentono di identificare e quindi reperire tutte le *Unità Informative*.
.idx[3)] **.vcb** **.ref**	File degli indici
.ths **.thv**[4)]	Files di Thesaurus, dove utilizzato
.tip **.tit**	Files contenenti la *cache* dei documenti utilizzata a fini applicativi

Oltre ai files indicati, un archivio presenta normalmente anche alcune directory che sono:

Directory	Contenuto
nomearchivio	La directory “omonima” è la collocazione naturale dei dati XML. In essa vengono creati, direttamente o in una struttura di directory regolata dalle configurazioni dell'archivio stesso, tutti i files XML che ospitano al loro interno le *Unità Informative*[5)].
***nomearchivio*.file**	Directory che contiene files *allegati* alle *Unità Informative*. Essi possono essere organizzati in directory numerate se effettivamente assegnati alle *Unità Informative* secondo uno stile “indiretto” oppure, se presenti direttamente nella directory indicata, sono *allegati* considerati in “Area di Parcheggio”. Essi sono in attesa di essere associati ad un'*Unità Informativa* o, se ciò non dovesse avvenire, di essere rimossi.
***nomearchivio*.chk**	Directory nella quale vengono registrati i files di `locking` delle *Unità Informative*. La regolazione sugli accessi di queste ultime, infatti, rappresenta una sorta di prenotazione con tanto di scadenza.

La Creazione dell'Archivio

Creare un archivio può significare cose piuttosto diverse.

Condizione	Descrizione	
Archivio inesistente	La creazione produce un archivio[6)] completo di mappa ed indici ma del tutto vuoto e privo di ogni file XML che rappresenti dei dati. Essi possono essere aggiungi in seguito con singoli inserimenti o in modo massivo.	
Archivio preesistente	**Creazione per Ricostruzione**	
Archivio preesistente	**Creazione per Ripristino**	La creazione in questo caso, consiste nell'azzeramento della mappa, degli indici e dei valori seriali precedentemente calcolati. I dati XML rimangono invariati ma la rimozione dei seriali comporta che una successiva ricostruzione dell'archivio non sarà del tutto equivalente all'originale e nuove operazioni sull'archivio stesso potranno generare, ad esempio, il “riempimento” dei buchi di numerazione lasciati sui campi seriali.
Archivio preesistente	**Creazione per Rinnovo**	La creazione per rinnovo comporta l'azzeramento di mappa, indici e seriali così come dei dati XML associati all'archivio. In questo modo l'archivio torna esattamente nella sua condizione originaria.

La creazione di un archivio deve combinare due componenti: un nome logico ed un percorso fisico in cui l'archivio viene creato.

Operando tramite Console[7] si può agire in diversi modi:

Azione	Risultato
Impostazione del nome logico dell'archivio nonché del suo nome reale	Si assume che il nome reale indicato sia quello che assumeranno tutti i files dell'archivio e che ad esso verrà aggiunta, file per file, l'estensione più appropriata. In sostanza quindi il nome espresso è un nome di file[8]. Qualora per quel nome logico sia già esistente un archivio si verificherà la piena congruenza tra il percorso dell'archivio esistente e quello indicato provocando errore se i due non combaciano perfettamente. Se, al contrario, non esiste alcun accoppiamento tra il nome logico ed un qualsiasi percorso, l'archivio verrà generato. In ambo i casi, in presenza di un file *nomearchivio*.conf.xml esso verrà comunque considerato valido e non verrà toccato in alcun modo mente in sua assenza esso verrà creato con contenuti d'esempio ed impostazioni di default.
Impostazione del nome logico dell'archivio e di un etichetta[9] come nome reale	Comportamento in tutto e per tutto analogo al precedente ma in tal caso il server calcolerà il nome fisico derivandolo da quello impostato. Ad esempio, se viene indicato il nome logico 'myarc' ed il percorso 'thearc' ed i programmi sono installati in 'c:\3di.it\extraway\xw\bin' il server creerà nella directory di default degli archivi[10] una directory col nome fisico indicato ad al suo interno gli archivi con lo stesso nome. Ciò sarebbe equivalente, in pratica, ad indicare come nome fisico 'c:\3di.it\extraway\xw\db\thearc\thearc'. L'archivio creato sarà corredato anche da una directory omonima che conterrà i dati.

Per ragioni pressoché storiche, la creazione/rigenerazione di un archivio effettuata tramite Console comporta la generazione di un archivio inesistente ovvero una 'Creazione per Ricostruzione' essendo essa la più frequente delle modalità di ricostruzione.
Ciò nonostante, dalla versione 23 del server, è consentito aggiungere al nome fisico dell'archivio dei modificatori che il server interpreterà come segue[11]:

Modificatore	Significato
<noxml>	Indica che i dati dell'archivio non troveranno posto in files XML bensì in un sistema di files binario (proprietario) della piattaforma eXtraWay. Particolarmente utile per la creazione, ad esempio, di archivi per CD rom quanlora non si intenda mantenere su di essi i dati in chiaro.
<reset>	Indica che si vuole ricostruire/rigenerare un archivio esistente. Se l'archivio non esiste la procedura da errore. Analogamente per creare un archivio inesistente tale parametro dev'essere assente.
<serials>	Indica, unitamente a '<reset>', che si vogliono salvaguardare i seriali già assegnati. E' la tipica configurazione da 'Creazione per Ricostruzione'.
<blind>	Modalità introdotta con il server di classe '23'. Consente di non dover specificare la natura della *Unità Informative* e quindi di stabilire una serie di comportamenti di default per inserire, modificare e ricercare. I documenti vengono quindi riconosciuti e salvati indipendentemente dalla presenza di configurazioni speciali nel file *nomearchivio*.conf.xml e con essi anche i *search_alias* che vengono a loro volta calcolati per default. In questo caso il file *nomearchivio*.conf.xml creato di default in caso di archivio inesistente ha un contenuto differente da quello di base e già adeguato allo scopo.
<drop>	Rappresenta la modalità necesaria per la 'Creazione per Rinnovo' in cui tutti i dati esistenti vanno persi assieme all'azzeramento di tutti i files che compongono l'archivio.

La Ricostruzione della Mappa

A meno che la Creazione effettuata non sia motivata da un'esigenza di Rinnovo, in tutti gli altri casi di intervento su un archivio esistente[12] la finalità della Creazione che azzera l'archivio è quella di proseguire con la ricostruzione dello stesso. Quello che può cambiare è il desiderio o meno di ricostruire integralmente l'archivio e salvaguardare o meno i seriali assegnati, ma questi sono aspetti già descritti in precedenza.
Assumendo quindi che, una volta azzerati mappa ed indici, si voglia procedere con la ricostruzione dell'archivio, esso passa da 3 fasi fondamentali[13]:

1. Rigenerazione della Mappa sulla base degli XML disponibili nell'archivio.
2. Rigenerazione degli indici delle *Unità Informative* presenti nella mappa ricostruita.
3. Rigenerazione della cache delle *Unità Informative* definita se pure non del tutto propriamente *Titoli*.

Queste operazioni possono essere compiute tramite Console in un solo *step*[14]. Vale comunque la pena di ricordare quanto segue:

- La ricostruzione della mappa, propedeutica alla altre fasi, si ottiene indicando il nome logico dell'archivio sul quale si intende operare ed il percorso nel quale si trovano i dati XML.
  Riprendendo l'esempio precedente, se il nome fisico di un archivio[15] è 'c:\3di.it\extraway\xw\db\thearc\thearc', il nome della directory nel quale si troveranno i dati XML dai quali ricostruire la mappa non potrà che essere 'c:\3di.it\extraway\xw\db\thearc\thearc\'.
  Indicando la stringa '.\' si sottindente la directory di default dell'archivio nella quale sono collocati i dati XML[16]
  A meno che non venga imposta[17] la verifica della *Wellformedness* dei dati dei quali si compie la mappa, questa fase è solitamente molto rapida.
- La ricostruzione della *Cache* dell'archivio è un'operazione piuttosto gravosa dal punto di vista computazionale ma di grande utilità applicativa. Nel ricostruire un archivio può essere importante renderlo disponibile il più rapidamente possibile ed in questo caso questa fase va saltata per essere eseguita in un secondo momento[18]
  Questa fase è solitamente molto più onerosa della semplice ricostruzione della Mappa.
- La ricostruzione degli indici è indispensabile per il corretto uso di un archivi, sia dal punto di vista della consultazione che della redazione[19].
  Essa dev'essere eseguita a seguito di una ricostruzione d'archivio ma può essere eseguita indipendentemente da essa. In

altre parole, quindi, qualora gli indici risultino danneggiati e sia suggeribile la loro ricostruzione, essa può (e ragionevolmente deve) essere effettuata senza ricostruire la Mappa[20] dell'archivio.
La rapidità di questa fase dipende in gran parte dalla configurazione del server e delle risorse[21] che vengono messe a sua disposizione.

1)
Fanno eccezione gli archivi nei quali si dichiari un repository di altro tipo, che sia un repository binario in forma proprietaria eXtraWay o un altro repository come ad esempio un data base Oracle.
2)
Solo per le versioni 23 e successive del server eXtraWay il file non è più un XMl ma un file binario la cui estensione diviene .stat
3)
Dalla versione 23 del server questo file si sdoppia in .idx.bpt e .idx.edt
4)
Quanto detto in precedenza per il file .idx che cambia dalla versione 23 del server vale anche per questi due files
5)
Scopo di quest'architettura è di rendere comprensibile la struttura data ai dati indipendentemente dal server eXtaWay e ricalcare in modo fisico l'organizzazione logica che si è volta dare ad essi
6)
Crea una directory con tutti i files necessari
7)
Consultare la documentazione ad essa relativa
8)
E non semplicemente un nome di directory
9)
Sovente lo stesso nome logico
10)
directory 'db' parallela a quella dei programmi
11)
L'aggiunta di questi modificatori anziché l'implementazione, ad esempio in Console, di chiamate differenti, affonda le radici nelle stesse ragioni storiche definite prima.
Di fatto è più pratico aggiungere modificatori “indolori” ad un comando che doverlo cambiare così da non incappare in incompatibilità col passato di componenti della piattaforma come il `Broker`
12)
Ma anche nel caso si disponga di un archivio nuovo al quale si vogliono imporre dato XML provenienti da altre fonti
13)
L'ordine di esecuzione delle stesse è palese: La prima fase è propedeutica alla realizzazione delle successive che, per contro, possono essere svolte in qualsiasi ordine in quanto indipendenti l'una dall'altra
14)
Si consulti la documentazione relativa per maggiori chiarimenti.
15)
estensioni escluse
16)
Ciò ci indica indirettamente che si può fare la mappa di dati contenuti in files XML collocati in posizioni diverse dalla directory di default dell'archivio.
Questa è una pratica lecita per quanto, di fatto, molto raramente utilizzata
17)
Tramite configurazione dell'archivio
18)
ATTENZIOE: dalla versione '22' del server eXtaWay, la gestione dei titoli è stata completamente riveduta. Ciò comporta che la ricostruzinoe massiva della *Cache* dell'archivio, fatte le opportune configurazioni, può essere eseguita senza effettivi e gravi impatti sul restantecomportamento del server.
19)
Dalla versione '22' del server è stata implementata una separazione tra indici *primari* e *secondari* che ha importanti impatti sulle performance di questa fase.
Vedasi la documentazione a ciò relativa.
20)
E conseguentemente anche la *Cache*.
21)
In particolare la RAM