

NICOLA FERRO - GIANMARIA SILVELLO  
*Dipartimento di Ingegneria dell'Informazione,  
Università degli Studi Padova*

### 1. Introduzione

Il Sistema Informativo Archivistico Regionale (SIAR) è un progetto della Regione del Veneto finalizzato allo sviluppo di un sistema di archivi digitale (*DAS – Digital Archive System*). Il principale obiettivo del SIAR è lo sviluppo di un sistema per redigere, gestire e condividere metadati di tipo archivistico in un ambiente distribuito. Il progetto SIAR è il risultato della sinergia tra competenze informatiche e archivistiche ed è volto a porre l'utente al centro del processo d'ideazione, progettazione e sviluppo del sistema: infatti, gli utenti sono stati coinvolti in ogni passo del percorso di progettazione e sviluppo del SIAR e il *feedback* raccolto è stato preso in considerazione in tutte le varie fasi del progetto. Si sono considerati i principi base dell'ingegneria del software per determinare sei fasi fondamentali che hanno caratterizzato e che caratterizzano il progetto SIAR:

*Ideazione:* in questa fase si sono definiti gli obiettivi e le priorità del progetto nonché le modalità con cui condurre il progetto; in particolare, questo ha significato individuare le modalità di lavoro comune tra informatici e archivisti. Si è preso dunque in considerazione la natura stessa degli archivi e delle descrizioni archivistiche. Inoltre, si è analizzato lo stato dell'arte dei sistemi di archivi digitali.

*Analisi dei requisiti:* in questa fase si è definito l'insieme di requisiti che il sistema SIAR deve soddisfare per incontrare le necessità di archivisti e altri utenti del sistema.

*Progettazione:* in questa fase si sono definite le funzionalità del sistema e le specifiche tecniche che ne permettono la realizzazione.

*Modello dati:* in questa fase si è definito un modello dati innovativo basato su un'organizzazione a insiemi annidati che permette di rappresentare, gestire e manipolare dati gerarchici in modo flessibile. Le funzionalità e le potenzialità di questo modello sono state discusse con gli archivisti così da valutare la possibilità di utilizzare tale modello a insiemi per modellare gli archivi e le descrizioni archivistiche.

*Sviluppo:* in questa fase si è sviluppato il sistema sia implementando il modello dati a insiemi sia sfruttando gli standard tecnologici utilizzati nel campo di ricerca delle biblioteche digitali verificando se queste fossero adattabili a un sistema di archivi digitali.

*Verifica di laboratorio:* in questa fase il sistema è stato testato e le varie funzionalità sono state verificate dagli archivisti. I suggerimenti e le critiche raccolte sono state un prezioso strumento per comprendere quali requisiti fossero soddisfatti dal sistema e dove si dovesse intervenire per migliorarlo.

## 2. Ideazione e stato dell'arte

La fase d'ideazione ha preso in considerazione la natura stessa degli archivi e delle descrizioni archivistiche. Il ruolo degli utenti "archivisti" è fondamentale per capire le caratteristiche e peculiarità degli archivi e quindi per risolvere i problemi che sorgono quando ci muoviamo dagli archivi fisici alle loro rappresentazioni digitali.

I documenti archivistici sono collegati tra loro e allo stesso tempo al loro ambiente di creazione e conservazione. Per conservare ed esprimere in modo completo il loro potere informativo nel tempo, gli archivi devono mantenere questa rete di relazioni. In particolare, le relazioni tra entità archivistiche sono di tipo gerarchico, dove ogni entità è correttamente identificata e interpretata esclusivamente in relazione all'entità che la contiene<sup>1</sup>. La soluzione principalmente adottata dai sistemi archivistici digitali è di rappresentare queste relazioni gerarchiche tra entità archivistiche

mediante una struttura ad albero.

In questo contesto, le descrizioni archivistiche sono lo strumento di maggiore importanza per gestire l'archivio. Le descrizioni archivistiche possono essere definite come "il processo di analisi, organizzazione e registrazione dei dettagli riguardanti gli elementi di un record o di una collezione di record, con l'obiettivo di facilitare il lavoro di identificazione, gestione e comprensione" e devono riflettere l'organizzazione e le caratteristiche dell'archivio.

Lo standard di descrizione archivistica in genere adottato in ambiente digitale è l'*Encoded Archival Description* (EAD). EAD è uno standard rilasciato dalla "Library of Congress" in unione con la "Society of American Archivists"<sup>2</sup>. Questo standard permette descrizioni multiple su più livelli, permette di rappresentare l'organizzazione gerarchica dell'archivio, di esprimere le relazioni fra le entità che lo compongono e di mantenere i loro legami con l'ambiente di creazione e conservazione. EAD è composto da una struttura flessibile e adattabile a diversi contesti applicativi e per tali ragioni è stato internazionalmente adottato da molti archivi<sup>3</sup>. Grazie a queste caratteristiche EAD consente di rappresentare e gestire le caratteristiche fondamentali dell'archivio anche in ambiente digitale.

D'altro canto EAD garantisce un'ampia libertà nella pratica di descrizione archivistica che è problematica in fase di elaborazione automatica dei file EAD. Infatti è spesso difficile sapere in anticipo come sono stati usati i diversi campi che lo compongono, ad esempio lo stesso campo descrittivo può essere interpretato diversamente da archivisti diversi e si può prestare a differenti utilizzi non sempre prevedibili. È difficile conoscere in anticipo come un determinato campo sarà utilizzato e quindi senza un'analisi manuale dei contenuti è difficile riuscire ad elaborare ed estrarre informazioni da file EAD provenienti da sorgenti diverse. Uno degli obiettivi principali per cui EAD è stato concepito è quello di favorire l'interoperabilità tra sistemi diversi, ma proprio l'eccessiva flessibilità del suo utilizzo rischia di essere una barriera in questo senso. Invero, com'è stato più volte riscontrato in letteratura, solamente i file EAD

che rispettano delle linee guida stringenti sono effettivamente condivisibili e ricercabili<sup>4</sup>. Inoltre, EAD non richiede una descrizione standardizzata del livello del materiale descritto e quindi diventa complicato accedere individualmente ad una particolare descrizione senza essere costretti a navigare l'intera struttura gerarchica dell'archivio. In ultima istanza, la ricerca e la condivisione di file EAD è resa difficile dalla profonda struttura gerarchica del file dal momento che un singolo file EAD descrive un intero archivio e non una singola entità archivistica; frequentemente, gli utenti sono interessati all'informazione che risiede a livello dell'unità che tipicamente è sepolta a un livello profondo della descrizione e quindi non è sempre facilmente raggiungibile.

EAD non consente di cambiare il formato di metadati usato per descrivere le entità archivistiche ed esprime insieme sia la struttura sia il contenuto delle descrizioni archivistiche. Questo fa sì che l'archivio sia gestito e trattato come un monolite impedendone una gestione flessibile e un'evoluzione (anche tecnologica) nel tempo<sup>5</sup>. Infine, l'utilizzo di EAD in molti casi costituisce una barriera al multilinguismo<sup>6</sup>, all'accesso con granularità variabile alle descrizioni archivistiche e all'adozione di standard tecnologici per l'interoperabilità largamente utilizzati nel settore delle biblioteche digitali come l'*Open Archive Initiative – Protocol for Metadata Harvesting* (OAI-PMH)<sup>7</sup>.

Nel progetto SIAR si sono presi in considerazione sia gli aspetti positivi che negativi della pratica di descrizione archivistica in ambiente digitale e di EAD, con l'obiettivo di rappresentare e gestire le caratteristiche fondamentali degli archivi e allo stesso tempo cercando di superare alcuni dei limiti sopra illustrati.

### 3. *Analisi dei requisiti*

Il SIAR deve tenere in considerazione le caratteristiche strutturali degli archivi e delle descrizioni archivistiche in modo da consentirne una corretta gestione e rappresentazione.

Il primo requisito individuato deriva direttamente dalla pratica archivistica: la struttura gerarchica in cui le entità archivistiche sono organizzate deve essere mantenuta, così come le relazioni tra loro in modo tale da mantenere il contesto in cui sono state create e conservate. In particolare, i formati di metadati scelti per la descrizione dovrebbero essere flessibili ed estendibili in modo da poter gestire l'eterogeneità delle risorse archivistiche.

Il secondo requisito di sistema è estendere e modificare i formati di metadati scelti per la descrizione. Inoltre, si deve considerare che un archivio non è composto solamente di documenti e descrizioni archivistiche, ma comprende anche organizzazioni, istituzioni, famiglie, persone e istituti di conservazione che contribuiscono alla creazione (soggetti produttori) e alla conservazione (soggetti conservatori) degli archivi. Ognuno di questi soggetti deve essere descritto mediante un adeguato formato di metadati.

Il terzo requisito è che il SIAR sia in grado di gestire diversi formati di metadati e di creare relazioni tra loro e le descrizioni archivistiche.

L'ultimo requisito è che l'accesso, la modifica e la cancellazione di una descrizione archivistica, sia consentito solamente a utenti autorizzati. Il SIAR deve essere anche in grado di condividere descrizioni archivistiche a diversi livelli di granularità e appartenenti a qualsiasi livello gerarchico senza dover necessariamente condividere l'intero archivio.

#### *4. Progettazione*

La fase di progettazione si concentra sui requisiti che un sistema informativo archivistico deve soddisfare e su come coniugare queste necessità con le tecnologie allo stato dell'arte come il protocollo OAI-PMH per lo scambio di metadati in ambiente distribuito.

In questa fase il lavoro congiunto di informatici e archivisti è stato fondamentale per trovare un punto di equilibrio tra le possibilità tecnologiche e le necessità archivistiche. Una parte consistente

del lavoro si è concentrata sulla definizione dei formati di metadati delle descrizioni archivistiche e dei soggetti produttori e conservatori. Dunque, per le descrizioni archivistiche si è progettato un formato di metadati estendibile basato sul catalogo italiano per le risorse digitali sviluppato nel contesto del Sistema Archivistico Nazionale (SAN). Questa scelta permette di utilizzare diversi formati di metadati e allo stesso tempo di esportare i metadati del SIAR nel SAN.

### 5. Modello dati

Il modello dati definito e utilizzato nel SIAR è basato sul modello NESTOR (*NEsted SeTs for Object hierArchies*)<sup>8</sup> che definisce dei modelli dati basati su insiemi annidati per la modellazione, rappresentazione, gestione, accesso, scambio e reperimento di risorse informative organizzate in strutture gerarchiche. NESTOR è formato da due modelli distinti ma complementari: il *Nested Sets Model* (NS-M) e l'*Inverse Nested Sets Model* (INS-M).

Nella figura sottostante è possibile vedere un albero e la sua rappresentazione grafica alternativa basata sui due modelli ad insiemi che compongono NESTOR.

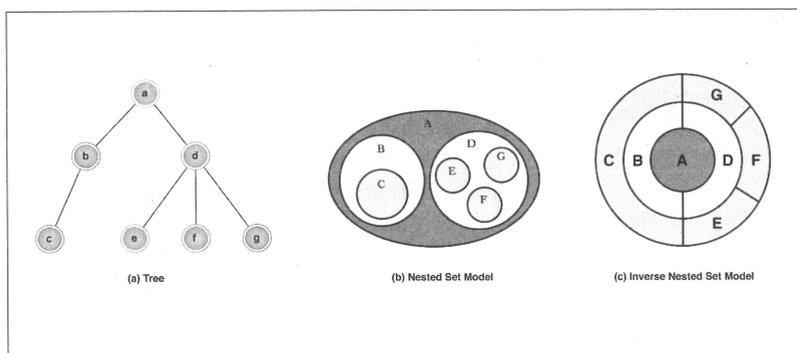


Figura 1 – (a) Un albero. (b) Una rappresentazione grafica del NS-M. (c) Una rappresentazione grafica dell'INS-M.

Il modello NS-M è rappresentato mediante un diagramma di Eulero-Venn, dove ogni nodo dell'albero in figura 1(a) è rappresentato da un insieme e la relazione gerarchica tra i nodi è espressa mediante la relazione di inclusione tra gli insiemi. Gli elementi appartenenti a ciascun insieme (non rappresentati in figura 1) rappresentano le risorse appartenenti a una specifica parte della gerarchia. Il secondo modello (INS-M) è basato sulla stessa idea del NS-M, ma ne inverte la logica. La rappresentazione dell'INS-M è basata sullo strumento grafico denominato "DocBall"<sup>9</sup>. Nell'INS-M ogni nodo dell'albero è rappresentato mediante un insieme che nella DocBall corrisponde a un settore circolare. L'INS-M definisce un insieme per la radice dell'albero (settoro circolare "A" in figura 1 (c)) e ogni insieme creato a partire dai nodi figli della radice è un soprainsieme di questo. Invece, nel NS-M tutti gli insiemi creati dai nodi figli della radice sono definiti come sottoinsiemi dell'insieme creato dalla radice dell'albero.

La differenza fondamentale di NESTOR con il modello ad albero risiede nella struttura matematica utilizzata per esprimere le relazioni tra gli elementi che compongono una gerarchia. L'albero è definito mediante l'utilizzo di relazioni binarie tra i nodi, mentre NESTOR esprime la gerarchia mediante la relazione d'inclusione tra insiemi.

I modelli dati presentati sono stati analizzati nel contesto del SIAR verificandone l'applicabilità nel caso degli archivi e si è verificato come l'adozione di questi modelli permetta di superare diversi limiti individuati nella fase d'ideazione. Grazie alla collaborazione con gli archivisti si è quindi analizzato come un archivio digitale possa essere rappresentato mediante il modello NESTOR. Se consideriamo un archivio costituito da diverse divisioni archivistiche dove ogni divisione contiene una serie di record, mediante i modelli ad insiemi è possibile modellare una divisione archivistica come un insieme e un record appartenente a questa divisione come un elemento appartenente all'insieme appena creato. In questo modo le relazioni tra divisioni archivistiche e le loro relazioni sono espresse mediante gli insiemi e le relazioni d'in-

clusione tra questi ed è possibile sfruttare la teoria degli insiemi per gestirli e manipolarli.

Il modello dati considerato non impone la scelta di un formato di metadati particolare e permette di separare chiaramente le parti strutturali da quelle di contenuto degli archivi. Queste caratteristiche aprono nuove possibilità per gli archivisti che possono progettare differenti schemi di metadati creati sulla base delle loro necessità, senza influenzare la struttura dell'archivio o il sistema archivistico digitale. Un altro vantaggio del modello NESTOR è che permette di separare la fase di modellazione da quella di scelta della tecnologia da adottare per la sua implementazione.

## 6. Sviluppo

L'architettura di sistema realizzata si divide in tre livelli principali. Il primo livello è costituito dall'infrastruttura di scambio dati<sup>10</sup>, il secondo dall'infrastruttura di gestione dei metadati<sup>11</sup> e il terzo dall'interfaccia utente.

L'architettura del SIAR si basa su un'implementazione del modello NESTOR che fa proprie alcune funzionalità native del protocollo OAI-PMH. In particolare, si sfruttano le funzioni di "selective harvesting" e l'organizzazione interna del protocollo basata su insiemi (*OAI-sets*) che possono essere utilizzati in maniera opportuna per implementare il modello NESTOR<sup>12</sup>. Di conseguenza, SIAR è l'esempio di un'integrazione profonda tra il modello dati scelto e il protocollo OAI-PMH; infatti, la gerarchia archivistica è mantenuta da un'organizzazione di OAI-Set e le descrizioni archivistiche sono modellate come metadati appartenenti all'insieme appropriato. Il sistema può gestire diversi formati di metadati e correntemente sono stati definiti degli schemi di metadati *ad hoc* per descrivere i rami alti dell'archivio, le unità archivistiche di diversa natura (ad esempio lettera e mappa), i soggetti produttori (uno schema diverso per enti, persone e famiglie), i soggetti conservatori e gli strumenti di ricerca. Il modello dati adottato permette

di cambiare il formato di metadati scelto senza interferire con la struttura dell'archivio o con il funzionamento del sistema. Questa funzionalità è già stata sfruttata durante lo sviluppo del SIAR, dal momento che i formati di metadati definiti sono stati rivisti e modificati in corso d'opera così da andare incontro alle nuove esigenze archivistiche e alle nuove linee guida nazionali definite nel contesto del SAN. Questa possibilità rende fruttuosa la collaborazione tra archivisti e informatici permettendo di riconsiderare le scelte progettuali senza influenzare in modo invasivo la fase di sviluppo del sistema. Oltre a ciò, la realizzazione del modello NESTOR sulla base del protocollo OAI-PMH consente di scambiare agevolmente i metadati archivistici con una granularità variabile e con la possibilità di ricostruire il contesto di una descrizione archivistica qualora necessario.

Il sistema SIAR prevede due tipi d'utente principale: l'utente archivistico e l'utente generico. Il primo può creare, modificare e cancellare i metadati, mentre il secondo può consultare i metadati inseriti nel sistema. La parte archivistica dell'interfaccia utente fornisce diverse moduli (*form*) per inserire e modificare i metadati. Si sono definiti diversi aiuti grafici per guidare gli archivisti nella redazione dei moduli d'inserimento e modifica dei metadati: ad esempio, le istruzioni su come compilare i campi di un modulo e una rappresentazione grafica ad albero degli archivi inseriti dove nuove divisioni archivistiche possono essere aggiunte alle esistenti agendo direttamente sui rami dell'albero; inoltre, si sono definiti diversi vocabolari controllati per guidare l'utente nel processo descrittivo.

Il sistema SIAR è sviluppato come un'applicazione *Web* il cui accesso avviene mediante un normale *Web Browser* e quindi è indipendente dal sistema operativo e dall'architettura hardware utilizzati. L'unico requisito di sistema per utilizzare il SIAR è disporre di una connessione a Internet; questa caratteristica è molto importante dal punto di vista della sostenibilità economica poiché permette anche a piccole organizzazioni distribuite sul territorio di utilizzare un sistema aperto per descrivere, gestire ed eventualmente condividere le proprie descrizioni archivistiche.

### 7. *Verifica di laboratorio*

La verifica del sistema con un gruppo selezionato di utenti esperti ha rappresentato (e tuttora rappresenta) un passo fondamentale nello sviluppo del SIAR perché permette di verificare il funzionamento delle funzionalità sviluppate e di testarne l'efficacia. La fase di verifica del sistema è strutturata in diversi incontri distribuiti nel tempo, con utenti archivisti chiamati a utilizzare il sistema e a segnalare eventuali malfunzionamenti o migliorie da apportare alle funzionalità fornite. Agli utenti è stato chiesto di inserire alcune descrizioni archivistiche riguardanti un archivio di loro competenza e dei metadati relativi a soggetti produttori e conservatori a esso collegati.

Le osservazioni degli utenti sono infine state raccolte e valutate dagli informatici e dagli archivisti che lavorano al SIAR e successivamente sono state apportate le modifiche necessarie al sistema, in modo da incontrare le richieste e le segnalazioni degli utenti.

La verifica di laboratorio è una fase importante per definire le priorità di sviluppo delle funzionalità del sistema e per delineare il programma delle attività future riguardanti il SIAR.

### 8. *Considerazioni conclusive*

Il progetto SIAR rappresenta un esempio rilevante di collaborazione tra informatici e archivisti finalizzato alla progettazione e sviluppo di un sistema archivistico digitale che tenga presente allo stesso tempo delle necessità archivistiche e delle potenzialità tecniche. Il costante processo di *feedback* svolto dagli utenti ha consentito di sviluppare un sistema *con* gli utenti e non solamente *per* gli utenti.

## NOTE

<sup>1</sup> S. VITALI, *Archival Information Systems in Italy and National Archival Portal* in M. AGOSTI, F. ESPOSITO, C. THANOS (a cura di), *IRCDL 2010. CCIS*, vol. 91, Sprienger, Heidelberg 2010, pp. 5-11.

<sup>2</sup> <http://www.loc.gov/ead/>

<sup>3</sup> K. KIESLING, *Metadata, Metadata, Everywhere - But Where is the Hook?*, *OCLC Systems and Services*, 17/2 (2001), pp. 84-88.

<sup>4</sup> C. J. PROM, C. A. RISHEL, S. W. SCHWARTZ, e K. J. FOX, *A Unified Platform for Archival Description and Access* in E. M. RASMUSSEN, R. R. LARSON, E. TOMS, e S. SUGIMOTO, (a cura di), *Proc. 7th ACM/IEEE Joint Conference on Digital Libraries*, (JCDL 2007), ACM Press, New York 2007, pp. 157-166.

<sup>5</sup> N. FERRO e G. SILVELLO, *A Methodology for Sharing Archival Descriptive Metadata in a Distributed Environment* in B. CHRISTENSEN-DALSGAARD et al. (a cura di), *Proc. 12th European Conference on Research and Advanced Technology for Digital Libraries* (ECDL 2008), *Lecture Notes in Computer Science* (LNCS) 5173, Springer, Heidelberg 2008, pp. 268-279.

<sup>6</sup> M. AGOSTI, N. FERRO, e G. SILVELLO, *Enabling Cross-Language Access to Archival Metadata in Cultural Heritage 2009: Empowering Users: An Active Role for User Communities*, Firenze University Press, Firenze 2010, pp. 179-183.

<sup>7</sup> N. FERRO e G. SILVELLO, *Design and Development of the Data Model of a Distributed DLS Architecture for Archive Metadata* in M. AGOSTI, F. ESPOSITO, e C. THANOS (a cura di), *5th IRCDL - Italian Research Conference on Digital Libraries*; *DELOS, An Association for Digital Libraries*, 2009, pp. 12-21.

<sup>8</sup> N. FERRO e G. SILVELLO, *The NESTOR Framework: How to Handle Hierarchical Data Structures*, *Research and Advanced Technology for Digital Libraries* (ECDL 2009), in *Lecture Notes in Computer Science* (LNCS) 5741 series, Springer-Verlag, pp. 215-226.

<sup>9</sup> J. VEGAS, F. CRESTANI, e P. DE LA FUENTE, *Context Representation for Web Search Results*, «*Journal of Information Science*», 33/1 (2007), pp. 77-94.

<sup>10</sup> N. FERRO e G. SILVELLO, *A Distributed Digital Library System Architecture for Archive Metadata*, 4th Italian Research Conference on Digital Libraries (IRCDL 2008), published by *DELOS, An Association for Digital Libraries*, 2008, pp. 99-104.

<sup>11</sup> Cfr. nota 7.

<sup>12</sup> M. AGOSTI, N. FERRO e G. SILVELLO, *The NESTOR Framework: Manage, Access and Exchange Hierarchical Data Structures*, *Proceedings of the 18th Italian Symposium on Advanced Database Systems*, Società Editrice Esculapio, Bologna 2010, pp. 242-253.