

Un utente, una password

FRANCO BUNGARO
ELENA MARANGONI

Sistema bibliotecario
Università di Torino
franco.bungaro@unito.it
elena.marangoni@unito.it

L'integrazione con il sistema di autenticazione d'ateneo di alcuni servizi bibliotecari nell'esperienza dell'Università di Torino

Il crescente interesse per il tema dell'identità digitale si accompagna al moltiplicarsi delle accezioni con cui tale espressione è utilizzata. Accezioni diverse che possono essere variamente collocate tra due estremi: da un lato una definizione essenzialmente teorico-filosofica, che fa riferimento alla presunta duplicità e ai rapporti fra identità reale e identità digitale, toccando temi quali la definizione del sé nell'era digitale, la diffusione di internet e dei social network, i rapporti tra mondi reali e virtuali; dall'altro una definizione prettamente tecnica, che ha piuttosto a che fare con temi quali i processi di autenticazione e autorizzazione e la loro sicurezza, i problemi connessi al trattamento dei dati personali, da armonizzare con l'esperienza dell'utente nell'uso di servizi online e gli obiettivi di snellimento e semplificazione delle procedure amministrative.¹

Per un'istituzione pubblica l'interesse per il tema è decisamente spostato verso questo secondo estremo e nella realtà dell'ateneo torinese si è concretizzato, a partire dal 2007, in iniziative promosse proprio dall'ambito bibliotecario.² In una fase di mutamenti sia organizzativi sia tecnici, coincisi tra l'altro con la nascita di un nuovo Polo SBN, obiettivi primari erano l'integrazione fra i servizi bibliotecari e gli altri servizi web d'ateneo e il miglioramento dell'usabilità dei servizi bibliotecari da parte degli utenti. Dal 2007 a oggi il Sistema bibliotecario ha raggiunto alcuni primi risultati, in particolare attraverso l'integrazione con il sistema di autenticazione d'ateneo di alcuni servizi, quali il deposito istituzionale ad accesso aperto "AperTO", il *virtual reference desk* "Preferito", il catalogo,³ utilizzando il software per il *single sign on* via web Shibboleth.⁴ Nei paragrafi che seguono si approfondisce in particolare quest'ultimo caso,

che ha portato anche alla messa a punto di un sistema di aggiornamento automatico delle anagrafiche utenti e di autenticazione tramite Shibboleth per l'accesso all'interfaccia di catalogazione di Sebina Open Library, realizzati in collaborazione fra il Sistema Bibliotecario d'ateneo, Data Management, il Cilea e le Divisioni servizi web integrati e sistemi informativi dell'Università di Torino.

[Prima fase: un nuovo Polo SBN per le biblioteche dell'Università di Torino.](#)
[Prima forma di integrazione delle anagrafiche](#)

Il Polo SBN dell'Università di Torino inizia la sua vita operativa il 27 febbraio del 2007. Per l'attività catalogafico-gestionale utilizza la versione v. 4.2 di Sebina, per l'OPAC poggia sul primo modulo della suite Open Library; le applicazioni sono ospitate su server del Cilea. Nato da una costola del Polo regionale piemontese il polo SBN delle biblioteche dell'Università di Torino è il cuore di un insieme di servizi digitali adottati all'interno di un unico progetto che mira a innovare e integrare l'offerta di servizi agli utenti istituzionali.

Innovazione, integrazione: la migrazione di centinaia di migliaia di dati⁵ da una base di dati ad un'altra⁶ è impresa che pone più di un problema da risolvere e che, in qualche caso, apre la porta a soluzioni mai sperimentate prima.⁷

Tra i record da migrare sono cruciali anche le liste degli utenti inseriti dalle biblioteche nell'applicativo usato per il prestito⁸ dal precedente sistema di gestione delle anagrafiche. L'export appare insoddisfacente e problematico (la base è regionale e comprende un significati-

vo quantitativo di utenti esterni, la forma delle intestazioni non è sempre coerente, i costi sono elevati in proporzione al beneficio sia in termini finanziari – sono previsti costi supplementari in uscita e in entrata – che di tempo – il periodo di chiusura del servizio si protrarrebbe per alcune ulteriori settimane), ma, in alternativa, è anche peggio costringere le biblioteche al reinserimento ex novo di tutti gli utenti nella nuova anagrafica, sia pure progressivamente.

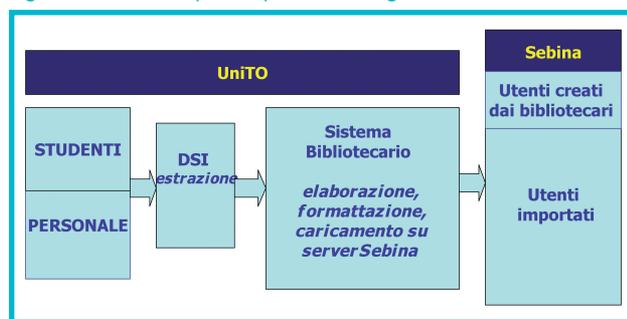
Oltretutto questa seconda ipotesi porterebbe a ricostruire manualmente per l'ennesima volta e con la ripetizione di errori e incongruenze una lista. Invece si sceglie un percorso innovativo realizzando, in via sperimentale, un'integrazione con i sistemi di gestione delle anagrafiche degli studenti e del personale dell'Ateneo, in modo tale da permettere successivi aggiornamenti incrementali dell'anagrafica utenti in Sebina. Il sistema – basato su una procedura per l'import dei dati messa a punto da Data Management – viene realizzato grazie alla collaborazione con la Divisione sistemi informativi e il Sistema portale, che forniscono le estrazioni dei dati anagrafici di studenti e personale di ateneo secondo un tracciato concordato, e lo Staff Programmazione Servizi Bibliografici Digitali del Sistema Bibliotecario che cura il controllo, la formattazione e la trasposizione dei dati in file di testo che vengono depositati sul server che ospita Sebina; da qui un'interrogazione in batch importa i dati nell'anagrafica Sebina, crea i nuovi utenti e aggiorna quelli già esistenti.

Con un primo import di oltre 100.000 utenti eseguito nel febbraio 2007, si inizia ad alimentare l'anagrafica del Polo Sbn, nella quale le biblioteche trovano già inseriti – ed è la prima volta! – gli utenti istituzionali (studenti e personale docente e tecnico amministrativo dell'Università di Torino).

Questo sistema rappresenta il primo passo nella direzione di quel principio di funzionamento delle procedure cui si tende in tutto l'ateneo: mettere a disposizione di tutti gli applicativi che li richiedono i dati anagrafici attingendo a un'unica fonte. L'aggiornamento di una e una sola tabella per ciascuna tipologia di utenti ricade a cascata su tutti i servizi beneficiari, abbandonando la creazione di anagrafiche parallele e parziali, generatrici di errori, ritardi nell'aggiornamento e disallineamenti fra le banche dati, che costringono alla ripetitiva compilazione di moduli destinati a chiedere all'infinito le stesse informazioni alle stesse persone.

Il sistema non tarda però a mettere in evidenza i suoi punti deboli. In primo luogo le procedure di controllo

Figura 1 – Workflow per l'import delle anagrafiche Unito in Sebina 4



e formattazione dei dati a carico del Sistema bibliotecario, che sono laboriose e complesse e rendono non soddisfacente la frequenza con cui è possibile aggiornare i dati (solo tre o quattro volte l'anno).⁹ In secondo luogo la scelta dell'identificativo univoco degli utenti, ovvero la matricola, poiché le persone che hanno più ruoli in ateneo (ad es. tecnici amministrativi che sono anche iscritti ad un corso di laurea) hanno due matricole e risultano duplicati in Sebina, con disagi per i bibliotecari in fase di registrazione dei prestiti.

Vi sono poi altri problemi ed esigenze quali la necessità di migliorare il livello di sicurezza nella trasmissione di dati¹⁰ e la mancata gestione delle cancellazioni; infatti la mancata comunicazione al Sistema bibliotecario degli studenti che lasciano l'Ateneo permette di eseguire solo import incrementali.

Parallelamente al proseguimento degli import, viene quindi progettato un nuovo sistema di aggiornamento delle anagrafiche utenti, maggiormente automatizzato, che possa ovviare ai problemi suaccennati.

[Seconda fase: autenticazione con Shibboleth e aggiornamenti dell'anagrafica](#)

La nuova procedura è entrata in produzione nel giugno 2010 dopo che il Polo UTO, cui aderiscono ormai tutte le biblioteche dell'Università, ha eseguito la migrazione al sistema Sebina Open Library (SOL) per tutti i moduli (catalogo, periodici, acquisti, circolazione documenti e gestione utenti). Nello stesso tempo viene introdotto un nuovo sistema di autenticazione, che permette di accedere all'ILS SOL con le credenziali di ateneo, il cosiddetto "sistema di credenziali unificato" – abbreviato in SCU – realizzato dal Sistema Portale per consentire l'accesso a tutti i servizi web a disposizione di studenti e personale.

Il nuovo login con credenziali SCU viene attivato sia per gli utenti che si vogliono autenticare in OPAC sia per i

bibliotecari che devono accedere al *back office* SOL, evitando di dover memorizzare username e password appositi per Sebina.

Il nuovo sistema di autenticazione ha come presupposti l'adesione dell'ateneo alla Federazione IDEM (IDentity Management per l'accesso federato)¹¹ e una stretta collaborazione tra la Divisione servizi web integrati, che mette a disposizione la tecnologia Shibboleth e le competenze necessarie alla realizzazione e mantenimento del nuovo sistema, il Sistema bibliotecario, che gestisce il servizio, Data Management, produttore dei software, e il Cilea, che amministra i server che ospitano SOL.

Shibboleth è il software utilizzato dalla federazione IDEM per gestire autenticazione e autorizzazione tra un utente, la sua organizzazione (l'Università di Torino) e un servizio web a cui l'utente deve accedere (in questo caso SOL). Permette un dialogo tra un identity provider dell'ateneo e un service provider installato da Data Management sul server SOL. L'utente inserisce le proprie credenziali SCU nella pagina di autenticazione, viene riconosciuto dall'identity provider dell'Università di Torino che ne trasmette l'autorizzazione e l'identità (definita grazie al codice fiscale) al service provider di SOL; SOL confronta il codice fiscale con le proprie anagrafiche di utenti e bibliotecari e permette l'accesso al sistema, secondo le sue abilitazioni.

Seconda fase: percorsi d'uso

I percorsi di autenticazione e autorizzazione sono due, distinti e differenti: uno destinato agli utenti del catalogo online <www.catalogo.unito.it>, l'altro per i bibliotecari che accedono alle funzioni di *back office* SOL.

L'utente che vuole accedere ai servizi personali del catalogo (per esempio per controllare la propria "situazione lettore", salvare ricerche e bibliografie, eseguire prenotazioni) una volta che ha cliccato sulle icone "login" o "spazio personale" viene condotto alla pagina di autenticazione dell'Università di Torino,¹² da dove, inserite le proprie credenziali SCU, viene riportato al sito del catalogo che lo ha riconosciuto quale utente autenticato attraverso una procedura del tutto trasparente.

La configurazione realizzata prevede che solo gli studenti e il personale dell'università possano eseguire il login in OPAC, escludendo ogni altro utente esterno. Tale scelta è stata motivata dal desiderio di offrire un servizio più semplice e intuitivo (uso delle sole credenziali unificate, accesso a una pagina di autenticazio-

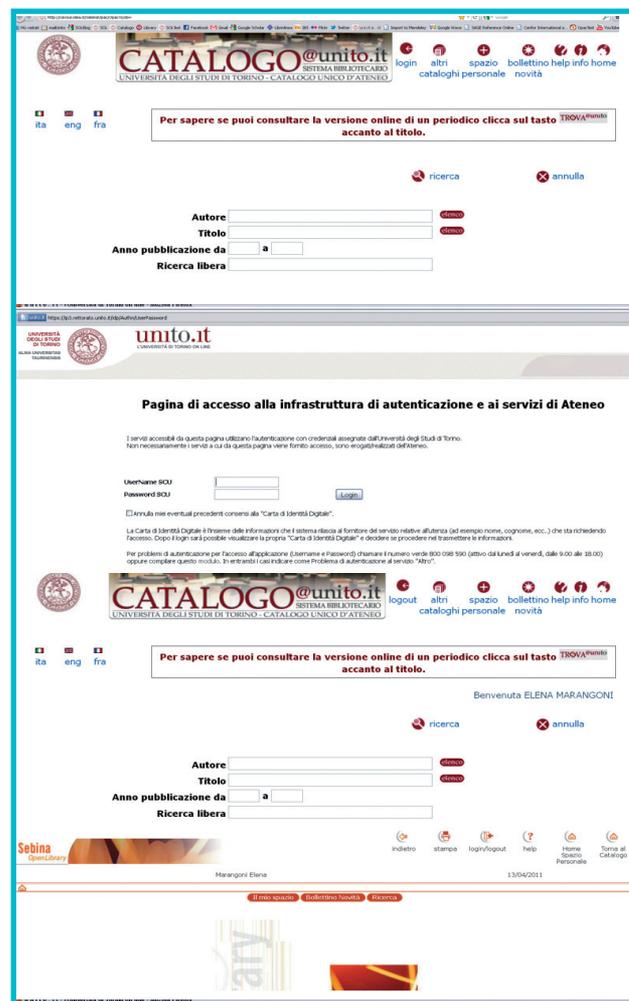
ne comune a quella di tutti gli altri servizi dell'ateneo che risulta perciò stesso familiare) per quanto limitato all'utenza istituzionale, in ogni caso di gran lunga la più interessata a un uso personalizzato del catalogo in linea.

D'altra parte la caratteristica pubblica dell'OPAC non viene certo tradita per questo. Tutte le funzioni tradizionali restano disponibili senza limitazioni.

L'accesso alle funzioni *back office* nell'ILS Sebina Open Library è possibile sia attraverso il login con credenziali SCU e Shibboleth, sia grazie al login con le credenziali create all'interno di SOL. Tale scelta è motivata dal fatto che nelle biblioteche si conta anche personale dipendente da cooperative esterne, talvolta in servizio per periodi molto brevi e privo di credenziali SCU.

Nella pagina iniziale, quindi, accanto al normale accesso con le credenziali SOL, è presente un collegamento al login Shibboleth. Scegliendo quest'ultimo il bi-

Figura 2 – Autenticazione in Catalogo e accesso allo spazio personale tramite le credenziali di ateneo (SCU)



bliotecario viene trasferito alla pagina di autenticazione dell'università e può inserire le proprie credenziali SCU. La richiesta di autenticazione è vagliata dall'identity provider Unito che trasmette a SOL il codice fiscale dell'utente. SOL esegue quindi un controllo nella propria anagrafica e, se il codice fiscale risulta associato a un profilo di bibliotecario, permette l'accesso al menù corrispondente alle abilitazioni di tale profilo, anche in questo caso attraverso passaggi automatici e del tutto trasparenti all'utente.

Si noti come il bibliotecario, a partire dal sito del *back office* dell'ILS <<http://uto.cilea.it/sebina>>, acceda, eseguendo l'autenticazione via Shibboleth, al menù SOL appropriato; mentre quando esegue l'autenticazione a partire dal sito dell'OPAC <www.catalogo.unito.it>, accede al menu dei servizi personali del catalogo, utilizzando sempre le medesime credenziali e seguendo un identico percorso di autenticazione.

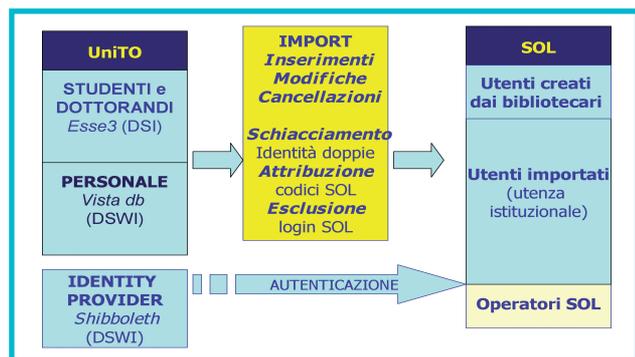
Seconda fase: procedura di aggiornamento delle anagrafiche

La figura 3 mostra la procedura di aggiornamento delle anagrafiche, alla base del funzionamento del login tramite Shibboleth giacché consente l'importazione in SOL del codice fiscale degli utenti, nonché gli altri dati utili per il modulo prestiti.

Le fonti dei dati importati in SOL sono due:

1) *il database di ESSE3*, il sistema del Cineca che l'Università di Torino ha adottato per la gestione delle anagrafiche studenti. È stata concordata l'estrazione settimanale di un file contenente i dati degli "studenti attivi". Tale file viene depositato su un server dell'ateneo, dove viene prelevato dalla procedura automatizzata di aggiornamento – che opera tramite interrogazione in batch – per eseguire l'import in SOL;

Figura 3 - Workflow per l'import delle anagrafiche Unito in SOL e l'autenticazione con Shibboleth



2) *il database di ateneo contenente le anagrafiche del personale docente e tecnico amministrativo*. È stato consentito al Sistema bibliotecario di accedere a una vista del database, dalla quale la procedura realizzata da Data Management estrae, ogni settimana per tutti record, una serie di campi relativi ai dati personali: indirizzo, telefono, struttura di appartenenza.

La trasmissione del file con le anagrafiche studenti tra Unito e SOL e l'interrogazione della vista del database delle anagrafiche del personale avvengono mediante una connessione VPN, che garantisce sicurezza nella trasmissione dei dati.¹³

L'estrazione da ESSE3 produce un file *differenziale*, che contiene anche le cancellazioni, permettendo l'eliminazione dei profili degli studenti che hanno terminato il corso di studi, sempre che non abbiano ancora movimenti in corso.

Invece, per le caratteristiche del database, non è possibile gestire le cancellazioni del personale; una carenza reputata accettabile dal momento che si tratta di un'utenza in generale assai più stabile di quella studentesca.

La procedura opera allora uno schiacciamento delle "identità doppie" confrontando i codici fiscali: qualora la stessa persona compaia due o più volte nelle due fonti dei dati (poiché ricopre più ruoli in ateneo ed ha quindi più di una matricola) la procedura importa comunque in SOL una sola anagrafica, corrispondente ad uno dei ruoli, scelto in base a una "gerarchia" decisa dai gestori di Polo. Tuttavia poiché le anagrafiche del personale vengono sovrascritte completamente ogni settimana, il ruolo presente nel database del personale prevale rispetto ad un altro eventualmente presente, per la stessa persona, nel database studenti.

La procedura poi, a partire da codici presenti nelle due fonti dei dati e mediante opportune tabelle di trascodifica compilate dai gestori di Polo, assegna a ciascun profilo utente da importare o aggiornare i codici "Tipo utente" (utilizzato per importare il codice relativo al ruolo ricoperto in ateneo, ad esempio: professore, ricercatore, studente, personale tecnico), "Tipo provenienza" (utilizzato per importare l'informazione relativa al corso di laurea cui lo studente è iscritto, il dipartimento di appartenenza del personale docente, o la struttura di lavoro del personale tecnico) e "Area di interesse" (usato per importare il codice relativo alla facoltà di appartenenza del personale docente). Gli altri dati importati sono: l'indirizzo, il telefono, il codice fiscale, l'indirizzo e-mail di ateneo.¹⁴

Infine, viene esclusa la possibilità di usare l'accesso con

credenziali SOL per tutti per gli utenti importati, che possono quindi usare il solo login Shibboleth.

Dopo tali rielaborazioni gli utenti vengono infine creati o aggiornati all'interno dell'anagrafica di SOL, dove vanno ad aggiungersi agli utenti creati dai bibliotecari.

È stata infatti mantenuta la possibilità di creare nuovi utenti da parte delle biblioteche, dal momento che una quota di utenza non è inclusa nell'utenza istituzionale (professionisti, studenti o docenti di altri atenei, studiosi non afferenti all'Università ecc.) o per supplire all'occorrenza a sempre possibili ritardi od omissioni nella trasmissione dei dati tramite la procedura automatizzata.

In ogni caso se gli utenti creati dalle biblioteche presentano un campo "codice fiscale" compilato correttamente, la procedura di import è in grado di riconoscere l'utente già esistente e di aggiornarne i dati, senza incorrere in duplicazioni.

Conclusioni

L'elaborazione e le prime sperimentazioni delle nuove procedure ci hanno insegnato che le strategie di integrazione dei servizi, accanto ai vantaggi, presentano quale inevitabile rovescio della medaglia una maggiore rigidità del sistema.

Infatti il corretto funzionamento delle procedure non può accontentarsi del solo accordo iniziale fra le diverse strutture dell'ateneo, ma pretende costanza di comunicazione; eventuali variazioni nelle fonti dei dati (come accade, per esempio, nel caso di modifiche dei codici di ruoli e corsi di laurea, o quando cambia la forma o l'estensione nel tracciato dei database) debbono ora essere comunicate anche al Sistema bibliotecario, pena l'immediata generazione di errori in SOL.

Più rigido e più consapevole diventa anche il comportamento dei bibliotecari che, dovendo ora autenticarsi con quelle stesse credenziali unificate che utilizzano ogni giorno per usare i più disparati servizi personali online, hanno completamente abbandonato la pratica di trasferire con leggerezza a terzi le proprie credenziali. Questo fatto, di per sé positivo, ha però costretto i gestori di Polo a rivedere le *policy* in uso sinora e ad adottare una gestione delle abilitazioni degli operatori più articolata e granulare, onde evitare che il nuovo sistema di autenticazione risulti penalizzante per il corretto lavoro dei bibliotecari.

Tuttavia ci pare che i vantaggi superino ampiamente i disagi che tali rigidità, almeno in un primo tempo,

comportano. Un primo riscontro oggettivo di questa considerazione viene dall'azzeramento quasi totale di richieste di assistenza relative all'autenticazione da parte di utenti dell'OPAC: l'uso delle credenziali unificate e il passaggio attraverso la familiare pagina di autenticazione sono di uso più facile per gli utenti, favorendo così il pieno utilizzo dei servizi offerti dal catalogo quali la possibilità di salvare ed esportare ricerche e bibliografie, controllare la propria situazione prestiti o eseguire prenotazioni.

Sul piano delle relazioni fra le strutture universitarie aver instaurato alcune prime forme di integrazione tra i servizi delle biblioteche e altri servizi appare un'esperienza da rafforzare in modo che le biblioteche possano continuare a partecipare allo sviluppo dei servizi web di ateneo nel loro complesso e ad esserne in parte protagoniste.

Tra i prossimi potenziali obiettivi si segnala l'accesso con *single sign-on* (ovvero una sola autenticazione effettuata per accedere ad uno qualunque tra i servizi web dell'università permetterà di accedere nel corso della stessa sessione a tutti gli altri senza dover ogni volta inserire login e password) o un migliore sfruttamento delle potenzialità di Shibboleth e della federazione IDEM per il riconoscimento e l'autorizzazione di utenti di altre istituzioni (come istituti di ricerca, altri atenei, aziende ospedaliere) la cui utenza è, o potrebbe diventare, in alcune circostanze utenza potenziale anche delle biblioteche dell'università.

NOTE

¹ Un esempio della varietà delle articolazioni del tema dell'identità digitale e dei diversissimi campi di interesse coinvolti è l'eterogeneo programma del convegno organizzato da CSI Piemonte "Le nuove ID.entità" (Torino, 17 novembre 2010) che ha visto confrontarsi filosofi ed esperti di comunicazione accanto a scienziati e tecnici informatici: <http://www.csipiemonte.it/convegni_scientifici/index.shtml>.

² Cfr. D. ARDUINO - F. BUNGARO - P. GARDOIS - A. SACCÀ, *L'autenticazione degli utenti per l'utilizzo dei servizi di biblioteca digitale*, intervento presentato al convegno "Autenticazione federata e biblioteche digitali" (Roma, 6 marzo 2007). Le slide dell'intervento sono consultabili al sito <http://www2.garr.it/aai2_slide/roma-2007-garr-presentation.pdf>.

³ I servizi di deposito istituzionale <<http://aperto.unito.it>> e il sistema informativo delle biblioteche dell'Università di Torino <<http://catalogo.unito.it>>, sono ospitati dal CILEA (le url effettive sono rispettivamente <<http://dSPACE-unito.cilea.it>> e <<http://cavour.cilea.it/SebinaOpac/Opac?sysb=>>>). Il servizio

di Virtuale Reference Desk è basato sui servizi interni dell'Università di Torino <<http://library.unito.it/preferito/>>.

⁴ <<http://shibboleth.internet2.edu>>.

⁵ La migrazione riguarda circa 600.000 descrizioni bibliografiche. A distanza di quattro anni (settembre 2011) i titoli compresi nella base dati UTO (questa la sigla dell'università di Torino in SBN) ammontano a circa 1.1 milioni.

⁶ La migrazione dal Polo regionale piemontese – che utilizzava a quel tempo un'applicativo basato sul DBMS Adabas – al Polo SBN dell'Università di Torino, è stata così realizzata: Il CSI Piemonte, presso cui era ospitato il server, ha consegnato lo scarico *industriale* al sistema bibliotecario dell'università che ha provveduto all'analisi e verifica dei dati e alla trasmissione a Data Management dove sono confluiti i dati bibliografici estratti in formato SBNMarc (allora ai suoi esordi) dai servizi dell'indice ICCU. DM ha infine provveduto al matching.

⁷ I record bibliografici vengono scaricati per la prima volta direttamente dall'indice in formato SBNMarc per un *merge* con i dati gestionali estratti tramite uno scarico *industriale* direttamente dalle tabelle del database.

⁸ Un'applicazione client server esterna al sistema catalografico e gestionale basata su *Informix*.

⁹ La rielaborazione consisteva nell'unificazione in una sola tabella delle informazioni contenute in file diversi, nella modifica dei codici (provenienza, tipo utente, area di interesse) per adattarli alle lunghezze dei campi accettate da Sebina, nelle formattazioni. Il risultato finale era un file testuale compilato secondo il tracciato messo a punto da Data Management per l'import.

¹⁰ Rispetto alla trasmissione via ftp che viene utilizzata in questa prima fase.

¹¹ La Federazione IDEM è un'infrastruttura di autorizzazione e autenticazione della rete Garr cui aderiscono 22 atenei italiani, oltre ad altri enti e partner esterni fornitori di servizi: cfr. <<http://www.idem.garr.it/>>. Sulla partecipazione dell'Università di Torino e i servizi resi disponibili agli utenti della federazione è disponibile la pagina <http://www.unito.it/unitoWAR/appmanager/istituzionale/speciali2?_nfls=false&_nfpb=true&_pageLabel=Speciali2_idem1>.

¹² La pagina di autenticazione, all'indirizzo <<https://ip3.rettora.to.unito.it/idp/Authn/UserPassword>>, è la pagina dell'identity provider Shibboleth di Unito (messo a disposizione dalla Divisione servizi web integrati). L'identity provider fornisce l'autenticazione confrontando username e password inserite dall'utente con l'LDAP d'ateneo. Quindi il riconoscimento dell'utente avviene tramite l'identity provider di Unito, che poi comunica a SOL un solo dato identificativo della persona, ovvero il codice fiscale, che diventa la chiave per il funzionamento di tutto il sistema.

¹³ Il tunnel VPN è stato installato tra un server dell'Università e il server che ospita SOL, presso Cilea. Per quanto riguarda invece le autorizzazioni al trattamento dei dati personali sono state rilasciate le necessarie nomine sia al Sistema Bibliotecario, sia alle ditte esterne coinvolte, Data Management e Cilea.

¹⁴ La scelta dell'email di ateneo come mezzo di comunicazione delle biblioteche con i lettori è stata presa sin dal 2007 e va nella direzione dell'integrazione delle biblioteche con gli altri servizi offerti a studenti e personale.

ABSTRACT

The article describes a case of integration between the catalogue and the personal data management systems realized at the University of Turin thanks to the collaboration between the Library System, other university divisions and Data Management, the company that provides the integrated library system Sebina. This process led in 2010 to the adoption of Shibboleth as the tool for the sign-in to the online catalogue and to the back office functions for the staff members. Students and teachers can now use the credentials provided by the university to have access to their catalogue personal pages. Furthermore automatic imports provides weekly updates of the database of libraries users.