



Hartwig Thomas, 4 dicembre 2013

Versione documento 1.20

SIARD Suite

Manuale



Edito da:

Archivio federale svizzero
Archivstrasse 24
3003 Berna
Svizzera

Prefazione

Il presente documento è un manuale tecnico dell'utente per l'applicazione *SIARD Suite* (Software-Independent Archival of Relational Databases) dell'archivio federale svizzero e descrive

- i presupposti tecnici per l'impiego
- l'installazione
- l'esecuzione

della *SIARD Suite* e delle sue componenti.

1 Indice

1	INDICE	3
2	INTRODUZIONE	4
2.1	DIRITTO D' AUTORE	4
3	CONCEZIONE SIARD	5
4	PRESUPPOSTI PER L'UTILIZZO DELLA <i>SIARD SUITE</i>	6
4.1	JAVA.....	6
4.2	BANCHE DATI	6
5	INSTALLAZIONE.....	8
6	COMPONENTI SIARD	9
7	SIARDEDIT.....	10
7.1	PRIMO AVVIO.....	10
7.2	FINESTRA PRINCIPALE	11
7.3	MENU	14
8	SIARDFROMDB.....	24
8.1	AVVIO.....	24
8.2	PARAMETRI D' AVVIO	24
8.3	OSSERVAZIONI	25
8.4	PREPARAZIONE DELLA BANCA DATI.....	25
9	SIARDTODB	30
9.1	AVVIO.....	30
9.2	PARAMETRI D' AVVIO	30
9.3	OSSERVAZIONI	31
9.4	PREPARAZIONE DELLA BANCA DATI.....	31
10	SISTEMI DI BANCHE DATI SUPPORTATI.....	35
11	FILE DI REGISTRO	36
12	LIMITAZIONI.....	37

2 Introduzione

L'archivio federale svizzero ha ideato, nell'ambito del progetto ARELDA (dal tedesco ARchivierung ELEktronischer DATen), l'applicazione SIARD (Software-Independent Archival of Relational Databases) per l'archiviazione a lungo termine dei contenuti di banche dati.

La Enter AG, su incarico dell'archivio federale, ha fatto esperienza pratica nel campo dell'archiviazione di banche dati produttive (2005, 2006) sviluppando infine una nuova versione di SIARD, nella quale sono messe in atto le conoscenze acquisite nella pratica per migliorare il prodotto.

Il presente documento è il manuale per la *SIARD Suite*.

2.1 Diritto d'autore

La *SIARD Suite* è un prodotto dell'archivio federale svizzero, presso il quale sono riservati tutti i diritti.

La *SIARD Suite* si basa sulla seguente componente di altri produttori:

- JavaHelp 2.0.05 (da <http://java.sun.com/products/javahelp>,
licenza: sulla cartella SIARD sotto *doc/JavaHelp-LICENSE.HTML*)

Questa componente è fornita con il CD SIARD come file JAR, per facilitare all'utente l'installazione. Gli utilizzatori della *SIARD Suite* sono tenuti a osservare le disposizioni di licenza contenute nella cartella */doc*.

3 Concezione SIARD

L'archivio federale svizzero è tenuto ad archiviare tutti i documenti della Confederazione, indipendentemente dal loro supporto informativo¹. Pertanto si doveva risolvere la questione dell'archiviazione a lungo termine di banche dati relazionali².

Nell'ambito del progetto ARELDA si è giunti alla conclusione che gli attuali sistemi di banche dati sono implementati in maniera troppo proprietaria, di modo che non è possibile un'archiviazione attendibile a lungo termine, basata solo su di essi.

Per garantire la disponibilità e l'accesso nel corso di lunghi periodi, è stata presa la decisione di memorizzare le banche dati in conformità a SQL:1999³, mentre i contenuti dei dati sono stati salvati in una collezione di file XML. Dato che il formato d'archivio per banche dati che ne risulta si basa su questi due standard ISO, si ritiene che la sua intelligibilità sia più duratura.

Il formato d'archivio per le banche dati è denominato formato SIARD.

Una caratteristica essenziale richiesta ai dati archiviati nel formato SIARD è che il contenuto della banca dati abbia „carattere di documento“, ovvero che i contenuti delle tabelle archiviate contengano i dati ufficiali dell'istituzione responsabile della banca dati, comprensibili indipendentemente dall'interfaccia usata (front end). Con la *SIARD Suite* non vengono archiviati né codici di programma eseguibili né oggetti, ma soltanto i dati di affari in tabelle database.

Nel formato SIARD la definizione dello schema di una banca dati archiviata è memorizzata in file XML conformi a SQL:1999, mentre la documentazione del contenuto delle tabelle e dei campi e i dati veri e propri sono registrati in forma di file XML. Per evitare che questi siano ampliati oltre misura, i dati BLOB e CLOB (Binary Large Objects e Character Large Objects) sono memorizzati in file (binari) separati, referenziati nei file XML.

In *SIARD Suite* vengono utilizzate da SQL:1999 soltanto le funzionalità principali, le „Core Features“ vere e proprie. Nelle versioni future della *SIARD Suite* il supporto potrà poi essere esteso ad altre componenti di SQL:1999, cosiddetti „Packages“.

In questo manuale non è ulteriormente spiegata la struttura del formato SIARD, giacché questa è descritta in un documento separato, fornito insieme alla *SIARD Suite*. Nel 2013 il formato SIARD è stato approvato quale standard eCH e può essere scaricato come eCH-0165⁴. Nella versione attuale di SIARD Suite il formato SIARD si distingue ancora dallo standard eCH per quanto riguarda le URL.

¹ Legge federale sull'archiviazione (Legge sull'archiviazione, LAr) del 26 giugno 1998 (Stato 1° agosto 2008)

² Archivierung von elektronischen digitalen Daten und Akten der Bundesverwaltung im Schweizerischen Bundesarchiv (ARELDA), presentazione progetto del 1.3.2001

³ Jim Melton, Alan R. Simon: *SQL:1999 – Understanding Relational Language Components*, Morgan Kaufmann Publishers, 2002, ISBN 1-55860-456-1 e

Jim Melton: *Advanced SQL:1999 – Understanding Object-Relational and Other Advanced Features*, Morgan Kaufmann Publishers, 2003, ISBN 1-55860-677-7

⁴ <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=it>

4 Presupposti per l'utilizzo della *SIARD Suite*

Per utilizzare la *SIARD Suite* è necessario che sia stata prima installata una versione JAVA. Presupposto per caricare o scaricare da o su un sistema di banca dati è un'infrastruttura corrispondente.

4.1 JAVA

Gli utilizzatori della *SIARD Suite* devono prima installare JAVA. Il presupposto tecnico minimo è JRE 1.6.

JAVA può essere scaricato liberamente dal sito <http://www.java.com/it/> (JRE – JAVA Runtime Environment) oppure <http://www.oracle.com/technetwork/java/javase/downloads/index.html> (JDK - JAVA Development Kit). Se non si è interessati a scaricare database di MS Access, si può utilizzare anche OpenJDK (<http://openjdk.java.net/>).

JAVA purtroppo, prima della versione 1.5, non ha dato alcuna possibilità di determinare la versione JAVA installata. Può, infatti, accadere che siano installate più versioni. Il cliente Oracle p.es. installa la propria versione JAVA (il più delle volte vecchia e difettosa) aggiungendola alle variabili di sistema PATH (!).

Si può verificare se la versione JAVA 1.6 o superiore sia disponibile o meno sotto Windows, avviando il registro di configurazione da *Start / Esegui* e digitando il comando *regedit*. Se JAVA 1.6 o superiore è installata, la chiave *CurrentVersion* in *HKEY_LOCAL_MACHINE\SOFTWARE\JavaSoft\Java Runtime Environment* ha il valore „1.6“ o superiore. Se il valore in registry non esiste o se è disponibile una *CurrentVersion* inferiore, allora dev'essere installata una nuova versione JAVA.

4.1.1 Architettura (32-bit/64-bit)

Dalla versione *JAVA 1.6* esiste la possibilità d'installare una versione di JAVA a 64-bit su un sistema operativo a 64-bit (p.es.: Windows 7, Ubuntu, MacIntosh), cosa consigliabile in generale. Il codice JAVA di *SIARD Suite* funziona sia sotto una versione a 32 bit sia sotto una versione a 64 bit di JAVA Runtime. Esiste tuttavia una dipendenza con l'architettura del sistema operativo per il ponte JDBC-ODBC, utilizzato per le banche dati di MS Access⁵. Se si utilizza la versione di JAVA a 64 bit, allora occorre che sia installato anche MS Access a 64 bit (o il relativo driver ACE ODBC). Con la versione JAVA a 64 bit non si può accedere a un'origine dati ODBC a 32 bit e con una versione JAVA a 32 bit non si può accedere a un'origine dati ODBC a 64 bit. Purtroppo il driver ODBC a 64 bit di Microsoft ha ancora parecchi bugs, cosa che rende impossibile la conversione di alcuni file di MS Access.

Se sono installate sia la versione a 32 bit sia quella a 64 bit di JAVA, *SIARD Suite* si avvia automaticamente con il Runtime a 64 bit. Si può comunque richiamare manualmente Runtime a 32 bit.

4.2 Banche dati

La *SIARD Suite* supporta al momento della sua prima fornitura i seguenti sistemi di gestione di banche dati:

- Oracle
- SQL Server
- MySQL (o MariaDB)

⁵ <http://www.enterag.ch/enterag/downloads/platform.xhtml>

SIARD Suite

- DB/2
- MS Access

In un secondo momento possono essere integrati altri sistemi di gestione di banche dati.

La *SIARD Suite* può accedere a queste banche dati con i seguenti driver JDBC:

sqljdbc4.jar	Solo per l'accesso ai database SQL Server 2000 fino al SQL Server 2012.
ojdbc6-fixed.jar	Solo per l'accesso ai database Oracle 8i fino a Oracle 11g (corregge errori nella distribuzione Oracle che impedisce la creazione di file XSD, mantenendo solo i packages Oracle).
mysql-connector-java-5.1.18-bin.jar	Solo per l'accesso ai database MySQL 5.5 e superiori.
sunjdbc.jar	Solo per l'accesso ai database MS Access via Open Database Connectivity di Microsoft ODBC 3.5
db2jcc.jar e db2jcc_license_cu.jar	Per l'accesso ai database DB/2 versione 8 e superiori.

Questi driver sono forniti e installati insieme a *SIARD Suite*. E' probabile che questi restino compatibili con versioni future di sistemi di gestione di banche dati ancora per alcuni anni.

5 Installazione

L'applicazione *SIARD Suite* può essere lanciata in ogni momento dalla chiavetta USB, dal CD-ROM o da un'unità di rete, sempre che sia disponibile una versione JAVA 1.6 o superiore installata correttamente, così da poter girare sotto Windows, LINUX, MacIntosh ecc. ecc.

Le impostazioni dell'utente sono salvate sotto `<home>/java/siard suite.properties`, dove `<home>` indica la cartella home dell'utente che dipende dalla piattaforma usata (sotto Windows XP questa è di solito `C:\Documenti e impostazioni\<userid>`, mentre sotto Windows 7 è `C:\utenti\<userid>`, .)

La *SIARD Suite* è fornita senza un programma d'installazione specifico, ma il programma principale *SiardEdit* verifica se, dove e in quale versione la *SIARD Suite* sia stata già installata dall'utente.

Per installare *SIARD Suite* quindi occorre semplicemente avviare *SiardEdit* – o con uno script specifico della piattaforma usata nella cartella principale (*siardedit.cmd* oppure *siardedit.sh*) o come applicazione JAVA lanciando `java -jar bin/SiardEdit.jar`.

Qualora *SiardEdit* costati che le impostazioni utente non siano ancora disponibili, allora ne propone l'installazione sul sistema dell'utente. La cartella standard a questo scopo è `<home>/applications/siard suite`.

SIARD Suite può disinstallata, reinstallata o installata altrove in ogni momento senza possedere i diritti di amministratore in *SiardEdit* al punto di menu *Strumenti*.

Anche se sul sistema è presente una vecchia versione della *SIARD Suite*, si può comunque lanciare, in parallelo, una nuova versione dalla chiavetta USB, dal CD-ROM o dall'unità di rete, senza modificare la vecchia versione.

Se la *SIARD Suite* viene installata sotto Windows, *SiardEdit* aggiunge automaticamente con l'aiuto dello script *desktop.cmd* un collegamento sul desktop e su *Start/Programmi*.



6 Componenti SIARD

La *SIARD Suite* è composta dai seguenti elementi:

- SiardEdit
- SiardFromDb
- SiardToDb

Il programma *SiardEdit* possiede un'interfaccia grafica utente interattivo che permette l'elaborazione di metadati in un archivio SIARD.

SiardEdit è lo strumento centrale per elaborare i dati in formato SIARD, mentre i dati primari non possono essere modificati. *SiardEdit* non è adatto per ricerche complesse. Per queste ultime si consiglia di caricare l'archivio SIARD in una banca dati e di compiere la ricerca con l'ausilio delle apposite tecniche usate nella gestione di banche dati.

L'applicazione per le righe di comando *SiardFromDb* serve a scaricare una banca dati e a memorizzarla in un file SIARD. La funzionalità di quest'applicazione è la stessa delle funzioni disponibili in *SiardEdit* al punto di menu *Crea un nuovo archivio da una banca dati* L'utilizzo della versione a righe di comando è più pratica, soprattutto per scaricare grandi banche dati e, grazie alla deviazione da *stdout* e *stderr*, lo scaricamento lento può essere meglio documentato (vedi „Utilizzo degli operatori di reindirizzamento dei comandi“ nella guida di Windows sotto *Start* oppure online⁶).

L'applicazione per le righe di comando *SiardToDb* serve al caricamento di una banca dati da un file SIARD. La funzionalità di quest'applicazione è la stessa delle funzioni disponibili in *SiardEdit* al punto di menu *Carica nella banca dati....* L'utilizzo della versione a righe di comando è più pratico, soprattutto per caricare grandi e, grazie alla deviazione da *stdout* e *stderr*, il caricamento lento può essere meglio documentato (vedi „Utilizzo degli operatori di reindirizzamento dei comandi“ nella guida di Windows sotto *Start* oppure online⁷).

⁶ <http://technet.microsoft.com/it-it/library/cc772622%28v=ws.10%29.aspx>

⁷ <http://technet.microsoft.com/it-it/library/cc772622%28v=ws.10%29.aspx>

7 SiardEdit

Il programma *SiardEdit* possiede un'interfaccia grafico utente interattivo che permette l'elaborazione di metadati in un archivio SIARD.

Con *SiardEdit* è possibile:

- creare un archivio SIARD „vuoto“ (senza dati primari) per ottenere una prima panoramica del processo d'archiviazione,
- confrontare i metadati di un archivio SIARD, che dev'essere generato, con l'ausilio dei metadati di un altro (vuoto o meno) archivio-modello (si riprendono tutte le descrizioni degli elementi che compaiono nell'archivio-modello),
- visualizzare i metadati di un archivio SIARD, ordinarli secondo diversi criteri, esaminarli, modificarli e completarli manualmente, sempre se i dati primari memorizzati nell'archivio SIARD non vengono toccati,
- visualizzare i dati primari di un archivio SIARD, ordinarli secondo diversi criteri ed esaminarli.

SiardEdit è lo strumento centrale per elaborare i dati in formato SIARD, mentre i dati primari non possono essere modificati. *SiardEdit* non è adatto per ricerche complesse. Per queste ultime si consiglia di caricare l'archivio SIARD in una banca dati e di compiere la ricerca con l'ausilio delle apposite tecniche usate nella gestione di banche dati.

La modifica di TIMES e TIMESTAMPS nella banca dati dipende dal fuso orario locale. Se per esempio si memorizza l'orario 15:30 a Zurigo, nel file XML è memorizzato l'orario UTC 14:30 (in inverno!). Se gli orari della banca dati devono essere interpretati invariabilmente come orari UTC, occorre avviare *SiardFromDb* con l'opzione

```
-Duser.timezone=GMT
```

Si può avviare *SiardEdit* con un file SIARD da aprire come unico argomento.

7.1 Primo avvio

La *SIARD Suite* è disponibile come pacchetto zip da scaricare. Il file va quindi decompresso ed estratto. Nella cartella *bin* del programma troverete il file *SiardEdit.jar*. Se JAVA è stato installato correttamente, il programma si avvia, sotto Windows, facendo un doppio clic su questo file. In alternativa si può avviare lo script specifico della piattaforma usata *siardedit.cmd* (per Windows) oppure *siardedit.sh* (per LINUX).

Nel caso ciò non funzionasse o se si lavora con un altro sistema operativo, *SiardEdit* può anche essere avviato dalla riga di comando nella cartella *bin* della *SIARD Suite* come segue:

```
javaw -jar SiardEdit.jar (Windows)
```

```
java -jar SiardEdit.jar (tutte le piattaforme)
```

Ciò è possibile a condizione che la cartella *bin* della versione JAVA si trovi in *PATH*, altrimenti occorre scrivere il percorso completo del programma eseguibile *java*.

Avviando *SiardEdit* per la prima volta, compare il messaggio seguente:

SIARD Suite



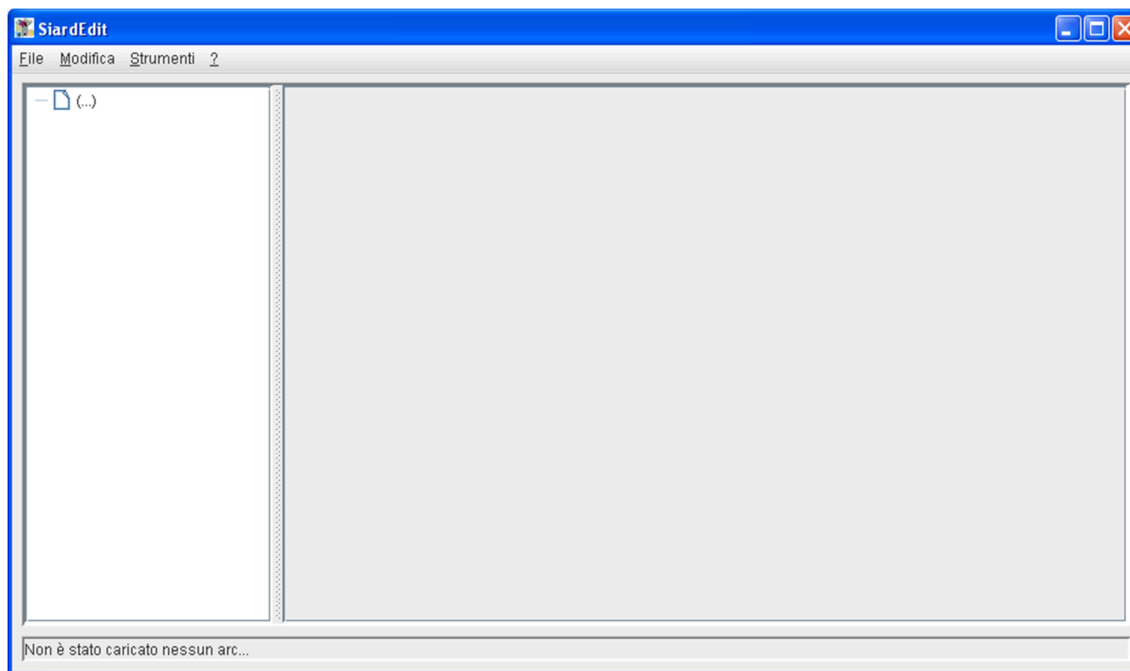
(Dato che *SiardEdit* ancora non conosce la lingua dell'utente, la lingua in cui appare questo messaggio dipende dalla lingua selezionata sul sistema operativo e al momento dell'installazione di JAVA.)

Rispondendo con *Sì*, si ha la possibilità d'indicare una cartella vuota o nuova, nella quale può essere collocata una copia di SIARD. Una volta installato, *SiardEdit* può essere poi avviato dalla cartella scelta.

Dopo l'avvio appare la finestra principale di *SiardEdit*, indipendentemente dal fatto che sia avviato dal CD-ROM o che sia stato installato sul PC dell'utente.

7.2 Finestra principale

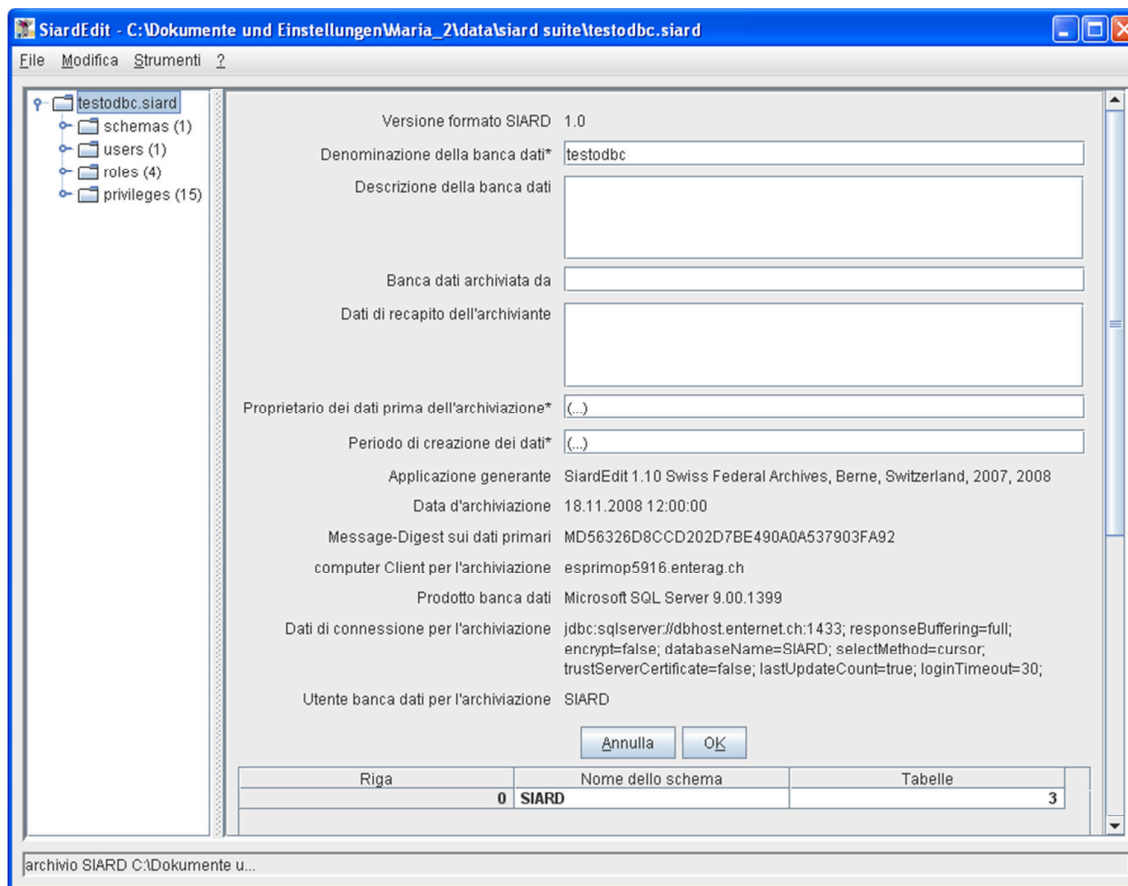
La finestra principale è composta dal menu (in alto), dall'albero di navigazione (a sinistra), dal contenuto (a destra) e dalla riga di stato (in basso).



SIARD Suite

Il bordo che divide l'area di navigazione da quella del contenuto, può essere spostato in modo flessibile, e anche l'estensione dell'intera finestra può essere dimensionata a piacimento fino alla grandezza minima.

Scaricando un file SIARD in *SiardEdit*, la finestra principale appare come segue:



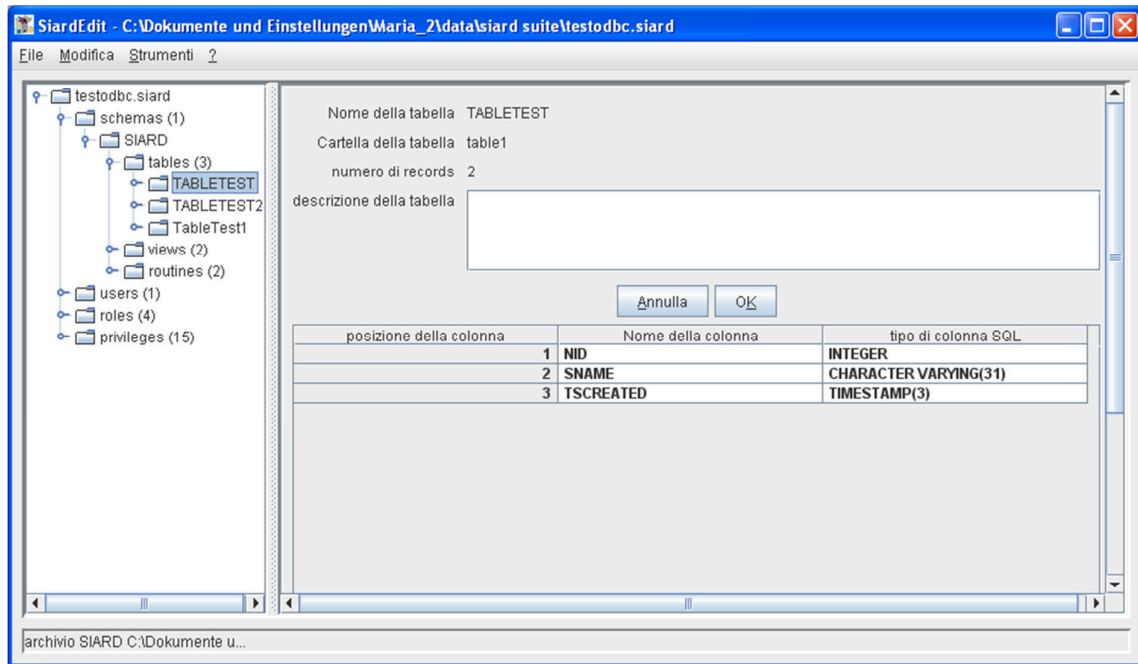
Nella sezione di sinistra si può navigare nell'albero dei metadati, mentre nella metà destra, in alto, si possono immettere o mutare i metadati modificabili che fanno parte della banca dati selezionata a sinistra.

Il pulsante *Riprendi* serve al trasferimento dei dati nei metadati del file SIARD attualmente aperto. Premendo *Annulla* invece, tutte le modifiche effettuate dall'ultima ripresa sono annullate.

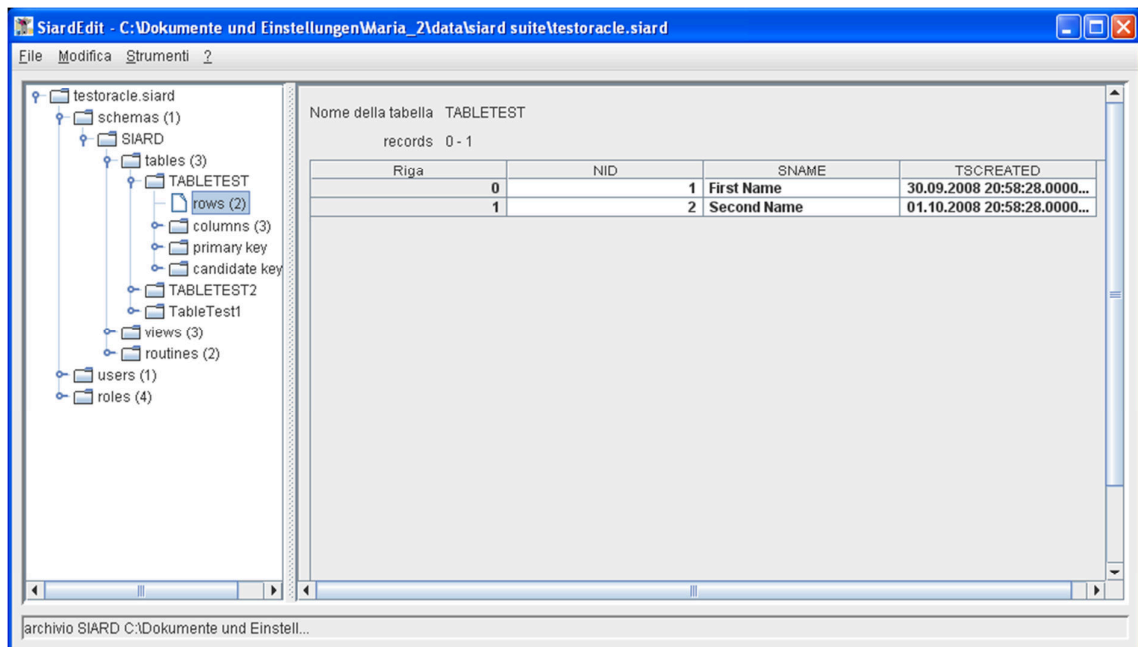
Sotto questi metadati compare una tabella dei sotto oggetti più importanti. Cliccando su un titolo di una colonna, la tabella può essere ordinata secondo questo titolo. Con un doppio clic su una riga, viene selezionato il sotto oggetto corrispondente.

Poiché le tabelle di uno schema o le colonne di una tabella nei metadati di per sé non hanno un ordinamento specifico, e vengono di solito disposte da *SiardEdit* in ordine alfabetico, questa funzione è molto utile per orientarsi nelle banche dati molto estese.

SIARD Suite



Navigando in quest'albero all'altezza della tabella, ci si può portare sui suoi dati primari, selezionando l'elemento *rows* situato immediatamente sotto.



SIARD Suite

Durante la visualizzazione dei dati primari, *SiardEdit* modifica leggermente il suo comportamento. Cliccando ora su un titolo di colonna, l'intera tabella viene ordinata secondo questa colonna, cosa che può durare a lungo nel caso di grandi tabelle!

Per consentire il lavoro anche con tabelle molto grandi, *SiardEdit* mostra solo ogni n -record (n è una potenza di 50), per evitare che debba essere visualizzata una finestra con un numero troppo grande di righe. Facendo un doppio clic su un numero di riga, vengono visualizzati i record contenuti tra questa riga e la successiva fino a raggiungere l'ultimo livello, dove tutti i record vengono visualizzati progressivamente.

Cliccando su una casella di testo, numerica o di data, il contenuto della casella viene visualizzato in un editor di testo. Facendo invece doppio clic su una casella binaria, il contenuto viene visualizzato in un editor binario. Nel menu *Strumenti / Opzioni* possono essere impostate le applicazioni da utilizzare per la visualizzazione delle caselle corrispondenti. Si può anche trascinare il bordo per ampliare le colonne, in modo da vedere completamente il contenuto delle colonne numeriche o di data.

Selezionando con un clic una riga della tabella, questa cambia di colore, mentre scegliendo il punto di menu *Modifica / copia (Ctrl-C)*, la riga può essere salvata negli appunti. Con il comando *Modifica / copia tutto (Ctrl-A)*, si può salvare tutta la tabella negli appunti. Il contenuto di quest'ultima può essere inserito in seguito, ad esempio, in un software foglio di calcolo per essere ulteriormente elaborato.

7.3 Menu

Nei sotto capitoli che seguono, sono descritti i singoli punti di menu.

7.3.1 File

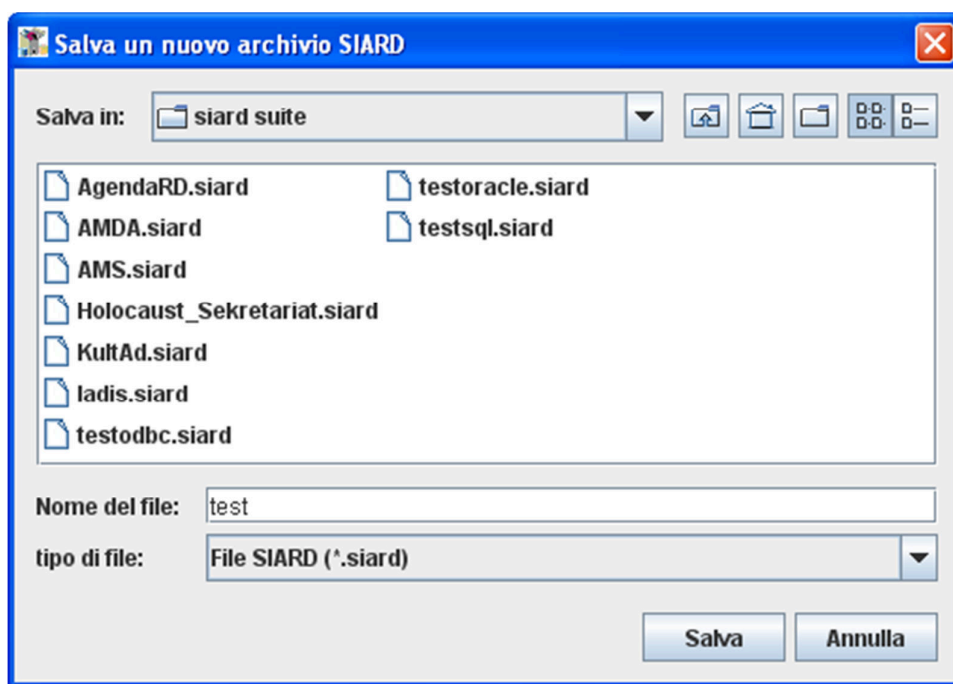
Crea un nuovo archivio da una banca dati...

Questo comando consente di scaricare una banca dati e di memorizzarla come file SIARD, che può essere poi visualizzato in *SiardEdit*.

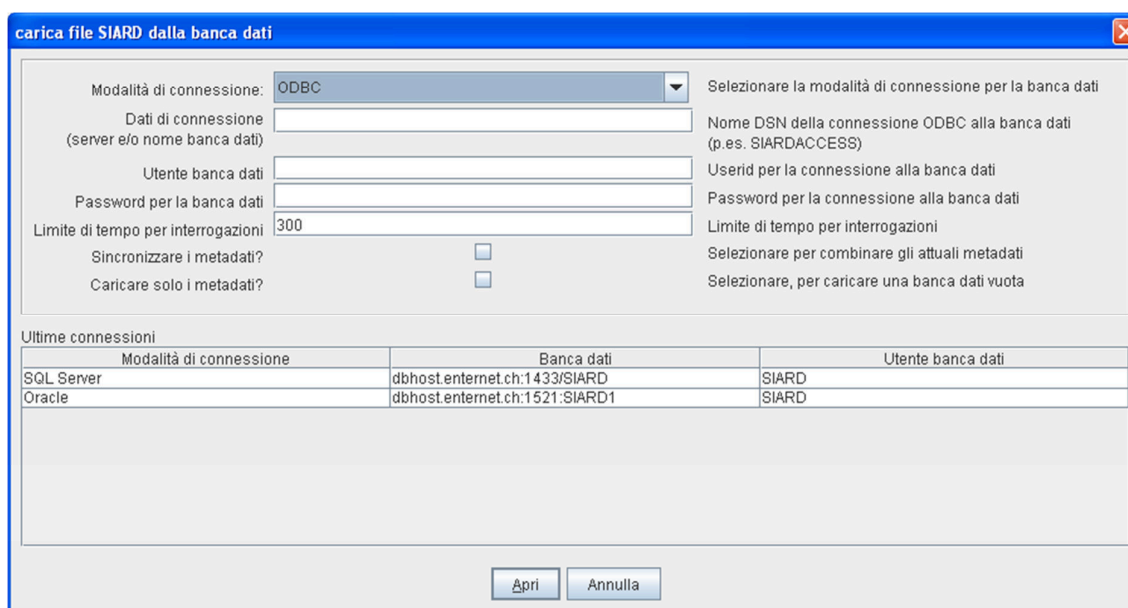
La funzione di questo comando è identica all'applicazione a riga di comando *SiardFromDb*. Nella descrizione di questo programma sono offerte informazioni importanti sulla preparazione dei diversi sistemi di banche dati a quest'operazione.

Selezionando questo comando la prima operazione richiesta è il nome sotto il quale salvare il nuovo file SIARD:

SIARD Suite



Nella finestra di dialogo successiva „Carica file SIARD dalla banca dati“ l'utente è quindi invitato a indicare i dati di connessione per la banca dati che dev'essere scaricata:



Modalità di connessione

Qui occorre selezionare la modalità di connessione

- *ODBC* per banche dati MS-Access,
- *Oracle* per banche dati Oracle

SIARD Suite

- *MySQL* per banche dati MySQL oppure
- *SQL Server* per banche dati SQL-Server
- *DB/2* per banche dati DB/2

Dati di connessione

I dati di connessione sono strutturati diversamente secondo la modalità di connessione.

ODBC

L'utilizzo della modalità di connessione *ODBC* (Open Database Connectivity) presuppone che sia stata impostata in precedenza una connessione ODBC per la banca dati desiderata. Per ulteriori informazioni si consulti la documentazione del sistema operativo in uso.

Nel campo Dati di connessione bisogna immettere il DSN (Datasource Name) della connessione ODBC.

Oracle

Dev'essere digitata una stringa di caratteri formata dai nomi DNS (DNS = Domain Name System) del server della banca dati, dal listenerport e dal SID (System ID) della banca dati, separati da doppi punti. Esempio per dati di connessione Oracle : *athlon700b.enterag.ch:1521:SIARD*.

In quest'esempio *athlon700b.enterag.ch* è il nome DNS del server, *1521* il nome del listenerport della banca dati (valore standard per database Oracle, sempre che la vostra configurazione non sia stata modificata dall'amministratore) e *SIARD* il SID della banca dati.

SQL Server

Qui occorre digitare una stringa di caratteri formata dai nomi DNS (DNS = Domain Name System) del server della banca dati, dal listenerport e dal nome della banca dati. Il server e il port sono separati da doppi punti, mentre il nome della banca dati è diviso da una barra. Omettendo il nome della banca dati, la connessione viene stabilita con la banca dati che l'utente ha indicato come accesso standard nel SQL Server.

Esempio per dati di connessione per un server SQL: *athlon800.enterag.ch:1433/siard*.

In quest'esempio „athlon800.enterag.ch“ è il nome DNS del server, *1433* il nome del listenerport della banca dati (valore standard per banche dati SQL Server, sempre che la vostra configurazione non sia stata modificata dall'amministratore) e *SIARD* il nome della banca dati.

MySQL

Qui occorre digitare una stringa di caratteri formata dai nomi DNS (DNS = Domain Name System) del server della banca dati, dal listenerport e dal nome della banca dati (=SchemaNamen). Il server e il port sono separati da doppi punti, mentre il nome della banca dati è diviso da una barra.

Esempio per dati di connessione MySQL: *asusp5b.enterag.ch:3306/SIARD*.

In quest'esempio „asusp5b.enterag.ch“ è il nome DNS del server, *3306* il nome del listenerport della banca dati (valore standard per database MySQL, sempre che la vostra configurazione non sia stata modificata dall'amministratore) e *SIARD* il nome del database.

DB/2

Qui occorre digitare una stringa di caratteri formata dai nomi DNS (DNS = Domain Name System) del server della banca dati, dal listenerport e dal nome della banca dati. Il server e il port sono separati da doppi punti, mentre il nome della banca dati è diviso da una barra.

Esempio per dati di connessione DB/2: *vmu64.enterag.ch:50000/SIARD*.

SIARD Suite

In quest'esempio „vmu64.enterag.ch“ è il nome DNS del server, 50000 il listenerport della banca dati (valore standard per banche dati DB/2 del tipo Express-C, sempre che la vostra configurazione non sia stata modificata dall'amministratore) e SIARD il nome del database.

Utente banca dati

Qui occorre digitare l'utente banca dati del database da archiviare. Attenzione: alcuni sistemi di banche dati ignorano la distinzione tra maiuscolo e minuscolo, altri invece no!

Se devono essere archiviate solo alcune parti del database, occorre allora creare un utente archivio assegnandogli unicamente dei diritti di lettura sugli oggetti da archiviare.

Password banca dati

Sia l'utente banca dati che la password del database sono generati dall'amministratore, il quale può richiederli nel caso non siano conosciuti. Attenzione: alcuni sistemi di banche dati ignorano la distinzione tra maiuscolo e minuscolo, altri invece no!

Nelle banche dati di MS Access l'utente si chiama di regola *Admin* e la password non è richiesta (riga di caratteri vuota).

Caricare solo i metadati?

Selezionando questa checkbox, viene generato un file SIARD senza dati primari. Ciò è molto utile se ci si vuole prima fare un'idea della grandezza stimata della banca dati. In questo file SIARD vuoto si possono già introdurre le descrizioni dei metadati e, più tardi, scaricando la banca dati vera e propria, aggiungerli a questa tramite la checkbox “sincronizzare”.

Sincronizzare?

Questa checkbox appare solo se nella finestra principale è stato già caricato un altro file SIARD. Selezionando quest'opzione, i metadati del file SIARD caricato in precedenza sono ripresi per quanto possibile nei metadati del nuovo file SIARD che è stato scaricato (sono cioè ripresi i dati riguardanti le tabelle, colonne ecc. che sono uguali in entrambi gli schemi)

Ultime connessioni

In questa tabella sono elencate le ultime connessioni effettuate dall'utente. Queste sono memorizzate nel file *siard.properties* nella cartella *java* situata nella cartella home dell'utente.

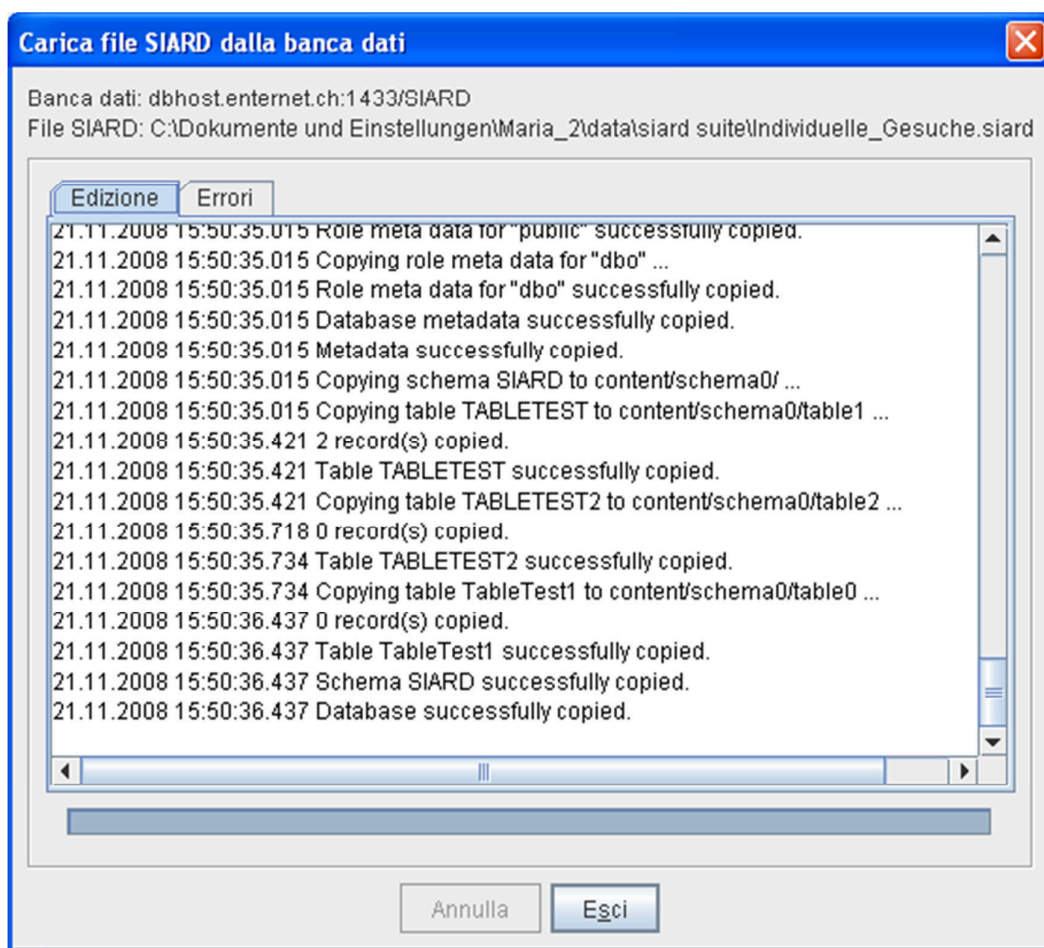
Cliccando una riga della tabella, i dati della connessione vengono completati. Solo la password resta vuota e il cursore salta sul campo password.

Apri

Con questo pulsante la banca dati viene caricata nel file SIARD selezionato e appare un protocollo identico a quello edito con *SiardFromDb*.

Interrompi

Con questo pulsante s'interrompe l'operazione.



Alla fine del caricamento, il protocollo può essere selezionato, copiato e stampato con l'ausilio di un programma di testo. Le indicazioni di tempo accanto ai singoli passi, servono a stimare la durata totale di caricamento di banche dati molto grandi.

Premendo qui il pulsante *Esci*, il nuovo file caricato compare aperto nella finestra principale.

Apri...

Con questo comando si apre un file SIARD (caricato in precedenza).

I vostri metadati dell'archivio SIARD compaiono quindi nella finestra principale e possono essere completati, modificati e salvati.

Salva (Ctrl-S)

Questo comando salva i metadati modificati nel file SIARD attualmente aperto.

Carica in una banca dati...

Questo comando serve a caricare il contenuto di un file SIARD aperto in un database.

La funzionalità di quest'azione è la stessa dell'applicazione per le righe di comando *SiardToDb*. Nella descrizione di *SiardToDb* sono contenute importanti informazioni su come preparare i diversi sistemi di banche dati per quest'operazione, come il fatto che occorra creare un utente banca dati idoneo con le autorizzazioni necessarie. Per caricare in MS Access bisogna allestire in precedenza una banca dati vuota e renderla accessibile tramite ODBC.

SIARD Suite

Selezionando questo comando, il dialogo *Carica file SIARD in una banca dati* invita l'utente a indicare i dati di connessione per la banca dati in cui bisogna caricare i dati.

carica file SIARD in una banca dati

Modalità di connessione: ODBC Selezione la modalità di connessione per la banca dati

Dati di connessione (server e/o nome banca dati) Nome DSN della connessione ODBC alla banca dati (p.es. SIARDACCESS)

Utente banca dati Userid per la connessione alla banca dati

Password per la banca dati Password per la connessione alla banca dati

Limite di tempo per interrogazioni 300 Limite di tempo per interrogazioni

Ultime connessioni

Modalità di connessione	Banca dati	Utente banca dati
SQL Server	dbhost.enternet.ch:1433/SIARD	SIARD
Oracle	dbhost.enternet.ch:1521:SIARD1	SIARD

Apri Annulla

Dopo che la connessione con la banca dati da caricare è avvenuta, *SiardEdit* visualizza una finestra di dialogo, mediante la quale si possono assegnare dei nomi agli schemi nell'archivio SIARD della banca dati.

Il caricamento dei contenuti di banche dati è soggetto alle limitazioni del sistema in cui i dati vengono caricati. Per esempio MS Access non supporta gli schemi delle banche dati, per cui in questo sistema tutte le tabelle sono caricate nella banca dati senza separazioni di schema.

Il vantaggio del caricamento di contenuti di banche dati consiste nel fatto che, in un sistema di banche dati, si ha a disposizione tutta la forza d'espressione di SQL per formulare le interrogazioni.

Esporta metadati...

Questo comando consente di esportare i metadati SIARD in formato XML.

Teoricamente si potrebbe anche aprire semplicemente un file SIARD con *PKZIP* e poi esportare i metadati dalla cartella *header*, però non tutti gli utenti hanno una licenza *PKZIP*. Questa sarebbe però necessaria, poiché altri programmi ZIP (Info-ZIP, cartelle ZIP di Microsoft Windows XP ecc.) non supportano (ancora?) il formato standard ZIP bit 64. Dato che i file SIARD possono diventare molto grandi, è stato necessario utilizzare questo formato bit 64 che consente di salvare dei file in formato ZIP più grandi di 4 GB e che contengono più di 66'000 file. Al posto di *PKZIP* si può utilizzare anche l'applicazione Open Source *Zip64File*⁸, per estrarre dati ZIP64.

Il file esportato *metadata.xml* può essere trasformato in un resoconto dei metadati della banca dati del file SIARD con l'ausilio di file di comando XSLT. Copiando il file *metadata.xsl* nella cartella *doc* dell'installazione SIARD nella stessa cartella in cui si trova il file esportato XML, aprendolo quindi in un browser più moderno, i metadati sono presentati in questo formato di resoconto.

⁸ <http://sourceforge.net/projects/zip64file/>

SIARD Suite

La descrizione della struttura dei metadati e del formato di file SIARD si trova nel documento *SIARD Format 2.0_en.pdf*, mentre la loro definizione formale nello schema XML *metadata.xsd* nella cartella *doc* dell'installazione SIARD.

Ultimi file utilizzati

Qui si trovano gli ultimi file aperti per accedervi rapidamente.

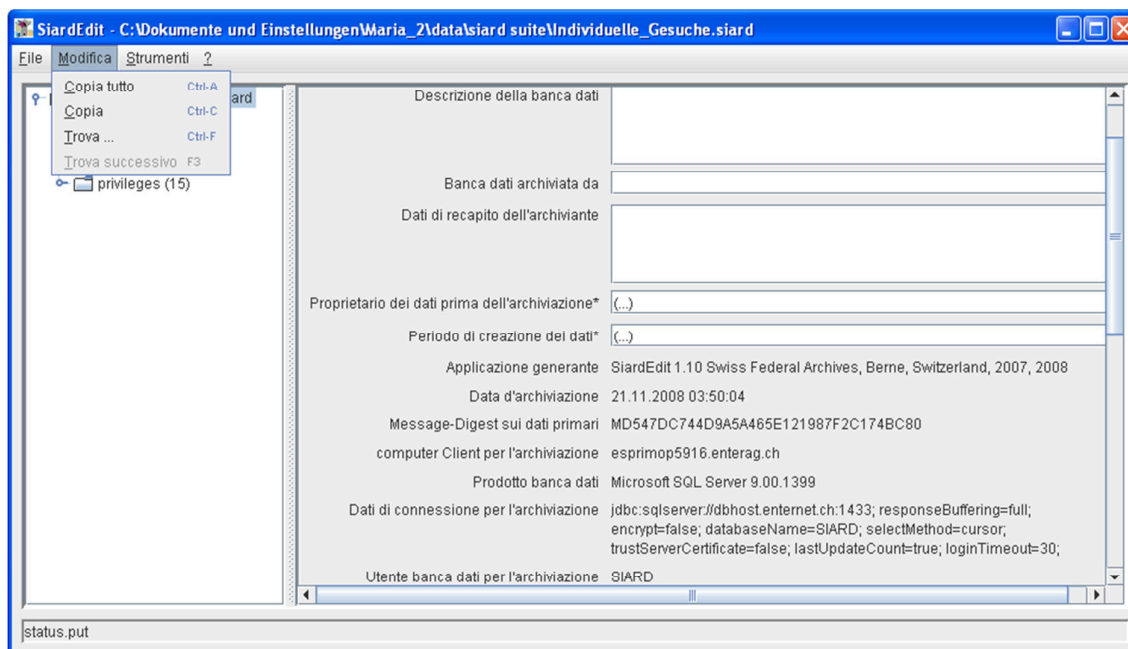
Esci

Nel menu *File* si trova *Esci* come ultimo comando per terminare l'applicazione.

7.3.2 Modifica

Copia tutto (Ctrl-A)

Nel menu *Modifica* troviamo il comando *Copia tutto (Ctrl-A)*.



Questo comando serve a copiare negli appunti la tabella visibile in basso a destra come testo separato da tabulatore, per poi poterla inserire per esempio in un programma di foglio di calcolo per la successiva elaborazione.

Copia (Ctrl-C)

Con questo comando si può riprendere la riga attuale, evidenziata in rosa, della tabella visibile in basso a destra come testo separato da tabulatore negli appunti, per poterla inserire in un'altra applicazione.

Trova... (Ctrl-F)

Questo comando consente la ricerca nei metadati di una stringa di caratteri qualsiasi.

Trova successivo (F3)

Con questo comando si possono trovare altri risultati della stringa di caratteri cercata.

SIARD Suite

7.3.3 Strumenti

Installa

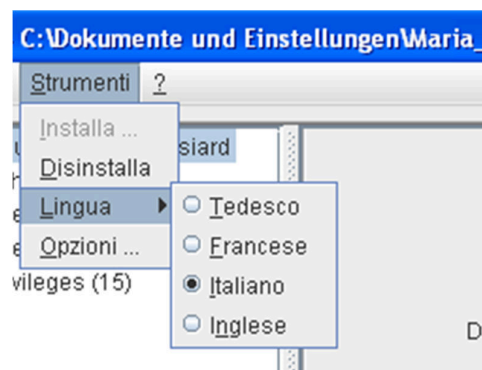
Nel menu *Strumenti* si trova anche il comando *Installa*, che consente l'installazione di SIARD sul PC dell'utente.

Disinstalla

Nel menu *Strumenti* si può rimuovere con questo comando l'installazione locale di SIARD.

Lingua

Nel menu *Strumenti* con il comando *Lingua* l'utente può selezionare la lingua desiderata.



La finestra principale di *SiardEdit* si adatta immediatamente alla lingua selezionata.

La lingua dell'utente viene memorizzata, terminando il programma, nel file `<home>/java/siard suite.properties`. Con `<home>` viene indicata sotto LINUX la cartella `$HOME`, e sotto Windows la cartella `C:\Documenti e impostazioni\<userid>`. (Per `<UserId>` s'intende il nome con il quale l'utente si è annunciato sotto Windows.)

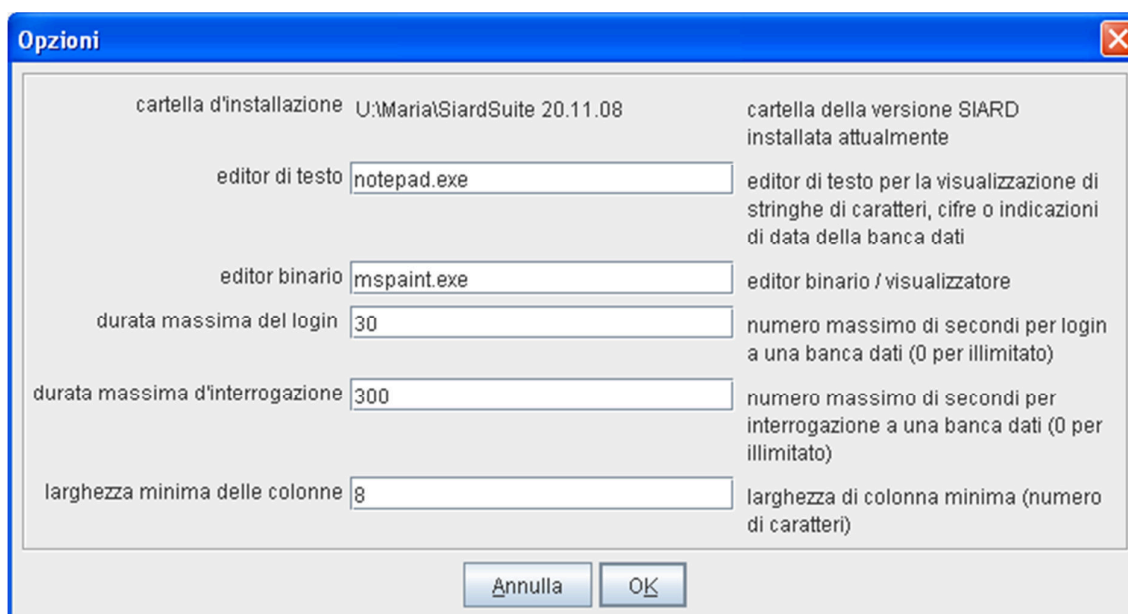
Utenti più esperti possono anche forzare il programma a usare un'altra cartella, scrivendo sulla riga di comando quando si avvia JAVA `"-Duser.dir=<altracartella>"`.

Esame dell'integrità

Il comando Esame dell'integrità nel menu Strumenti consente di verificare che i dati primari non siano stati modificati e che la loro sigla (MD5 o SHA-1) non sia stata calcolata e memorizzata nei metadati. *SIARD Suite* "sigilla" i dati primari solo una volta con questa sigla immediatamente dopo il download da una banca dati.

Opzioni ...

Nel menu Strumenti si trova il comando Opzioni, con il quale è possibile definire alcune impostazioni individuali:



I dati dell'*editor di testo* e dell'*editor binario* sono utilizzati se si fa un doppio clic su una cella di dati primari del tipo stringa di caratteri oppure del tipo dati binari. Il contenuto della cella viene salvato in un file temporaneo e aperto con l'applicazione indicata. L'editor binario dev'essere adattato di volta in volta secondo il tipo di dati CLOB.

La *durata massima d'annuncio* stabilisce dopo quanti secondi un tentativo di annunciarsi a una banca dati dev'essere interrotto. Questa durata dev'essere modificata soltanto se si ritiene che la rete o la banca dati reagiscano molto lentamente.

La *durata massima d'interrogazione* dovrebbe essere messa a zero per scaricare banche dati molto grandi, poiché in questo caso le interrogazioni possono durare più dei 5 minuti previsti nell'impostazione di base.

La *larghezza di colonna minima* definisce la rappresentazione della tabella nella parte inferiore destra della finestra.

7.3.4 ?

Guida ...

Questo comando visualizza l'indice della guida.

Info ...

Nel menu ? si trovano le informazioni sui diritti d'autore per *SiardEdit*:



Inoltre é visualizzata qui l'architettura del sistema operativo del runtime di JAVA.

7.3.5 Guida contestuale

L'utente ha in tutto il programma la possibilità di usufruire della guida contestuale, che è attivata premendo il pulsante F1. La guida visualizzata mostra un argomento attinente all'elemento o alla finestra interessata dal mouse o dalla tastiera.

Selezionando il comando Guida nel menu ?, viene richiamata la spiegazione della funzione di guida.

Nel menu della guida le singole pagine possono essere richiamate direttamente dall'indice ed è anche possibile compiere una ricerca per voci. Una pagina visualizzata viene sincronizzata di volta in volta con l'indice.

8 SiardFromDb

Il programma *SiardFromDb* è un'applicazione a righe di comando, che estrae una banca dati in un archivio SIARD. Con *SiardFromDb* si può da una banca dati

- Allestire un archivio SIARD „pieno“ (metadati e dati primari), basato sulla banca dati (opzione "full"),
- Allestire un archivio SIARD „vuoto“ (senza dati primari) con una definizione di schema basata sulla banca dati (opzione "empty").

8.1 Avvio

```
java -cp <siardpath>/bin/SiardEdit.jar ch.admin.bar.siard.SiardFromDb [-h] | [-e]
  [-c <config file>] -d <data source> -n <database name> -u <database user>
  -p <database password> -s <siard file> [-q <query timeout>] [-t <template file>]
```

Sotto <siardpath> occorre indicare la cartella nella quale SIARD è stato installato. Nella sottocartella *bin* della cartella d'installazione, si trova il file *SiardEdit.jar* con la classe *ch.admin.bar.siard.SiardFromDb*, il cui *main()* viene qua avviato con il comando *java* (sotto Windows è meglio utilizzare *javaw*).

Se c'è l'opzione *-h (help)* sulla riga di comando, viene visualizzata la sintassi d'avvio.

Dato che l'avvio contiene molti parametri, lo si può semplificare creando un file di configurazione nel quale questi vengono memorizzati, ad eccezione della password. Se questo file esiste, l'avvio si riduce a:

```
java -cp <siardpath>/SiardEdit.jar ch.admin.bar.siard.SiardFromDb -c <config file>
  -p <database password>
```

Se sono stati indicati tutti i parametri e in più un file di configurazione, i parametri nel file di configurazione sono completati da quelli sulla riga di comando. Ciò semplifica la ripetizione del download di SIARD.

Se è stata selezionata l'opzione *-e (empty)*, vengono archiviati solo i metadati.

8.2 Parametri d'avvio

<i>parametro</i>	<i>significato</i>
<data source>	ODBC, Oracle, "SQL Server", MySQL o "DB/2"
<database name>	Nome della banca dati, strutturata diversamente secondo la connessione: ODBC: Data source name (DSN) configurato della connessione ODBC Oracle: p.es. dbhost.enternet.ch:1521:SIARD1 SQL Server: p.es. dbhost.enternet.ch:1433/SIARD SQL Server: p.es. dbhost.enternet.ch:1433/SIARD MySQL: p.es. dbhost.enternet.ch:3306/SIARD

<i>parametro</i>	<i>significato</i>
	DB/2: p.es. dbhost.enternet.ch:50000/SIARD
<database user>	UserId per la sessione di download
<database password>	Password per la sessione di download
<siard file>	Nome di file dell'archivio SIARD da generare
<query timeout>	Timeout per interrogazioni SQL, indicato in secondi (o illimitato)
<template file>	File SIARD da usare come modello, per compilare manualmente i campi da riempire

8.3 Osservazioni

E' consigliabile eseguire il programma *SiardFromDb* su un „snapshot“ della banca dati, che non si modifica durante l'archiviazione.

O l'archiviazione è effettuata interamente con successo, oppure fallisce interamente.

Nel caso si abbia a che fare con grandi tabelle, si consiglia di scaricare in precedenza un file SIARD vuoto, in modo da ottenere una visione d'insieme di tutti i metadati e dell'estensione della tabella. Ciò aiuta a stimare la durata del download. Inoltre, nel caso di grandi tabelle, si dovrebbe selezionare l'opzione -q 0, poiché non si può prevedere quanti secondi duri un'interrogazione su una grande tabella.

La modifica di TIMES e TIMESTAMPS nella banca dati dipende dal fuso orario locale. Se per esempio si memorizza l'orario 15:30 a Zurigo, nel file XML è memorizzato l'orario UTC 14:30 (in inverno!). Se gli orari della banca dati devono essere interpretati invariabilmente come orari UTC, occorre avviare *SiardFromDb* con l'opzione

```
-Duser.timezone=GMT
```

8.4 Preparazione della banca dati

Secondo il tipo, la banca dati ha bisogno di una diversa preparazione per l'archiviazione. Dev'essere chiaro che la scelta dell'UserId per la banca dati utilizzata nel download, è decisiva per la grandezza della banca dati da archiviare. Il file SIARD che ne risulta conterrà gli oggetti del database visibili all'utente. Per questo motivo è sconsigliabile utilizzare un amministratore del database per il download (*SYS* o *SYSTEM* per Oracle, *dbo* per SQL Server, *mysql* o *schemi d'informazioni* per MySQL, *db2inst1* per DB/2 Express-C), poiché questi utenti hanno di regola troppi diritti di lettura per oggetti che non devono essere archiviati. Per le banche dati MS Access invece è consigliabile l'archiviazione di tutto il contenuto della banca dati, per tanto qui si raccomanda l'utilizzo di *Admin* come UserId della banca dati.

Il fatto che la scelta dell'UserId della banca dati determini la grandezza della banca dati archiviata, può essere utilizzato per controllare esattamente la grandezza dei dati che giungono nei file SIARD: si crea un nuovo utente d'archiviazione nella banca dati e gli si assegnano i diritti di lettura esattamente per gli oggetti che devono essere archiviati.

8.4.1 Preparazione MS Access

La gran parte delle banche dati MS Access può essere aperta con l'UserId *Admin* e la password „“. Per ottenere l'accesso a banche dati meglio protette, occorre consultare la documentazione del progettista.

SIARD Suite

File Accdb e Mdb

MS Access 2007 supporta due formati: il „vecchio“ formato dei file con l'estensione *.mdb* e un „nuovo“ formato con l'estensione *.accdb*. I file Accdb supportano nuovi tipi di file – in particolare campi con più valori e i nuovi tipi di file Allegato, anche a più valori. I file accdb proibiscono inoltre l'accesso a tabelle di sistema e rendono quindi impossibile l'accesso ai metadati ivi memorizzati tramite l'interfaccia ODBC (e quindi anche JDBC). Per i file Mdb si poteva impostare l'accesso alle tabelle di sistema *MSysObjects*, *MSysRelationships* e *MSysQueries*, cosa che consentiva l'accesso di *SIARD Suite* ad altri metadati, non accessibili tramite il *DatabaseMetaData* della classe JDBC.

SIARD Suite tratta la nuova situazione come segue: nel caso si possa accedere alle tabelle *MSys*, l'informazione supplementare viene memorizzata nell'archivio SIARD. Altrimenti non possono essere memorizzate né chiavi esterne né views. Per l'archiviazione questa perdita di metainformazioni si può trascurare, poiché non sono toccati dati primari.

Più problematico è invece il nuovo tipo di file *Allegato*. Microsoft ha stabilito che il suo contenuto in nessun caso può essere letto con l'interfaccia ODBC (e quindi JDBC). La versione attuale del formato SIARD non potrebbe neanche archiviare in modo appropriato tali campi BLOB a più valori. Per questo motivo *SIARD Suite* riprende tali campi – come fanno gli strumenti di migrazione di Microsoft – nel download solo come campi di testo che contengono liste di nomi di file degli allegati. Il contenuto degli allegati va comunque perso! Purtroppo Microsoft impedisce a *SIARD Suite* – e a tutti gli altri clients ODBC o JDBC –, di localizzare questa perdita. Rimane quindi all'archivista di archiviare separatamente gli allegati nei file Accdb.

Se occorre archiviare dei file Mdb, l'archivista ha la scelta di memorizzare solo quei metadati che sarebbero archiviati in un file accdb – senza chiave esterna e views –, o di preparare il file Mdb in modo che le tabelle di sistema siano accessibili per l'utente *Admin*.

Questioni di architettura (32-bit / 64-bit)

Se *SIARD Suite* è stata avviata con un Runtime JAVA a 32 bit, si può accedere solo a un driver ODBC a 32 bit. Se MS Access è installato su una piattaforma Windows a 64 bit, occorre impostare solo accessi ODBC a 64 bit con l'amministratore standard ODBC (*C:\Windows\system32\odbcad32.exe*). Questi sono quindi accessibili a *SIARD Suite*, soltanto se è stata avviata con un JAVA Runtime a 64 bit. Volendo allestire un accesso ODBC a 32 bit, occorre usare l'amministratore ODBC a 32 bit (*C:\Windows\SysWOW64\odbcad32.exe*). Questi accessi ODBC a 32 bit sono accessibili per *SIARD Suite*, soltanto se è stata avviata con un JAVA Runtime a 32 bit.

I file batch *siardedit.cmd*, *siardfrommdb.cmd* e *siardtodb.cmd* (e lo Shortcut impostato sul desktop dall'installazione) utilizzano JAVA Runtime a 64 bit, se questo è stato installato. Se allo stesso tempo è disponibile un JAVA Runtime a 32 bit, si può avviare manualmente *SIARD Suite* con quest'ultimo, richiamando i file batch con il parametro facoltativo */32*. Se sorgono problemi caricando un file MS Access con ODBC a 64 bit, conviene provare con ODBC a 32 bit. Attualmente il driver ODBC a 64 bit ha ancora troppi bugs.

Rendere accessibili le tabelle di sistema di file mdb

1. Per rendere accessibili all'utente di archivio (di solito *Admin*) le tabelle di sistema *MSysObjects*, *MSysRelationships* e *MSysQueries*.

MS Access prima del 2007:

Rendere visibili gli oggetti di sistema per l'utente *Admin* sotto *Strumenti / opzioni / visualizzazione*, e quindi sotto *Strumenti / protezione / autorizzazioni di gruppo e utente* autorizzare l'accesso alle tabelle *MSysObjects*, *MSysRelationships* e *MSysQueries*.

SIARD Suite

MS Access 2007:

Sotto *pulsante Microsoft Office / Opzioni di Access / Database corrente / Opzioni dell'applicazione... / Visualizza oggetti del database* si rendono visibili gli oggetti del database e quindi sotto *Strumenti database / Amministrazione / Utenti e autorizzazioni / Autorizzazioni gruppi e utenti*, rendere accessibili le tabelle *MSysObjects*, *MSysRelationships* e *MSysQueries* per l'utente *Admin*. (tutto il gruppo *Strumenti database / Amministrazione / Utenti e autorizzazioni / Autorizzazioni gruppi e utenti* manca, se il file aperto è un file accdb.)

2. Impostare una connessione ODBC, che fa riferimento al file MDB (vedi anche la voce ODBC nella guida di Windows sotto *Start / Guida di Windows* per consultare le istruzioni di come creare una connessione ODBC sotto *Start / pannello di controllo / strumenti di amministrazione / origine dati (ODBC)*), e nelle proprietà avanzate (sotto *avanzate ...*) e impostare *ExtendedAnsiSQL=1*. La connessione ODBC può essere impostata da un non-amministratore come *utente DSN*.

Configurazione automatica e conversione

Uno script open source *accesstosiard.js*⁹, sviluppato come complemento per *SIARD Suite*, esegue i lavori di configurazione per l'utente e converte un file ACCDB o MDB in un file SIARD.

8.4.2 Preparazione Oracle

La soluzione migliore per preparare un database Oracle è di configurarlo per l'esportazione come amministratore, creando un nuovo utente e dotandolo di tutte le autorizzazioni d'accesso agli oggetti da esportare. Con l'ausilio di questo nuovo utente, verificare in base ad interrogazioni modello se tutti i riferimenti rilevanti sono visualizzati. Si consiglia di commentare e documentare le interrogazioni modello per i futuri utenti dell'archivio.

Troubleshooting Oracle

Affinché SIARD riesca a stabilire una connessione con il database Oracle desiderato, occorre che le autorizzazioni, i firewall ecc. siano impostati in modo tale, che la connessione possa essere stabilita con il seguente comando:

```
sqlplus <database user>/<database password>@(description=(address=(protocol=TCP)
(host=<hostname>) (port=<port>)) (connect_data=(service_name=<service>)))
```

Quindi occorre utilizzare in SIARD Suite il nome del database *<hostname>:<port>:<service>*.

8.4.3 Preparazione SQL Server

La soluzione migliore per preparare un database SQL-Server è di configurarlo per l'esportazione come amministratore, creando un nuovo utente e dotandolo di tutti i permessi d'accesso agli oggetti da esportare.

Se non si accordano all'utente le facoltà del proprietario per gli schemi da archiviare, questi non potrà leggere i privilegi correnti e le informazioni sugli utenti e sui ruoli potranno essere incomplete.

Se l'utente riceve autorizzazioni SELECT, viene sempre archiviato automaticamente lo schema „dbo“ – cosa di solito indesiderata. Le impostazioni migliori per l'utente sono quindi:

⁹ <http://sourceforge.net/projects/accesstosiard/>

SIARD Suite

Accordare sotto permissions database solo autorizzazioni CONNECT. Sotto proprietà utente definire l'utente come proprietario di tutti gli schemi da archiviare. (attenzione: tutti i proprietari precedenti vengono „espropriati“. Eventualmente occorre ripristinare le impostazioni precedenti dopo l'archiviazione.)

Troubleshooting SQL Server

Affinché SIARD riesca a stabilire una connessione con il database SQL Server desiderato, occorre che le autorizzazioni, i firewall ecc. siano impostati in modo tale, che la connessione possa essere stabilita con il seguente comando:

```
sqlcmd -U <database user> -P <database password> -S tcp:<hostname>,<port> -d <database>
```

Quindi occorre utilizzare in SIARD Suite il nome del database *<hostname>:<port>/<database>*.

8.4.4 Preparazione MySQL

La soluzione migliore per preparare un database MySQL è di configurarlo per l'esportazione come amministratore, creando un nuovo utente e dotandolo di tutti i permessi d'accesso agli oggetti da esportare.

```
CREATE USER '<USER>' IDENTIFIED BY '<PASSWORD>';  
GRANT SELECT ON <SCHEMA>.* TO '<USER>';
```

(MySQL tratta database e schemi come termini intercambiabili. Al posto di <SCHEMA> occorre inserire qui per esempio il nome del database.)

Con l'ausilio di questo nuovo utente, verificare in base ad interrogazioni modello se tutti i riferimenti rilevanti sono visualizzati. Si consiglia di commentare e documentare le interrogazioni modello per i futuri utenti dell'archivio.

Troubleshooting MySQL

Affinché SIARD riesca a stabilire una connessione con il database MySQL desiderato, occorre che le autorizzazioni, i firewall ecc. siano impostati in modo tale, che la connessione possa essere stabilita con il seguente comando:

```
mysql -h<hostname> -P<port> -D<database> -u<database user> -p<database password>
```

Quindi occorre utilizzare in SIARD Suite il nome del database. (Attenzione: nessuno spazio tra -p e <password>!)

Per comprendere le autorizzazioni di un utente di dati MySQL, occorre considerare quattro tabelle in INFORMATION_SCHEMA: USER_PRIVILEGES (autorizzazioni globali), SCHEMA_PRIVILEGES (autorizzazioni agli schemi), TABLE_PRIVILEGES (autorizzazioni per le tabelle) e COLUMN_PRIVILEGES (autorizzazioni per le colonne).

8.4.5 Preparazione DB/2

DB/2 riconosce come utenti solo gli utenti OS. Bisogna quindi impostare in OS un utente adeguato e assegnare a questo tutte le autorizzazioni per il database da archiviare con GRANT ON DATABASE.

Attenzione: sul sistema operativo LINUX, l'UserId può essere composto solo da lettere minuscole e può essere lungo al massimo otto caratteri!

SIARD Suite

Troubleshooting DB/2

Sulla stazione di lavoro occorre che sia installato un client DB/2. Sotto Windows occorre quindi avviare *db2cmd*, per avere accesso alle righe di comando *db2*. Sotto LINUX si ha invece accesso direttamente.

Con le righe di comando *db2* si deve quindi impostare un nodo TCP/IP tramite CATALOG e infine stabilire una connessione con CONNECT.

A questo scopo bisogna avviare la finestra con le righe di comando per DB/2 o il command line processor (clp) per db2 e digitare quanto segue:

a) LIST NODE DIRECTORY

Questa directory dovrebbe elencare i node registrati sul client. Nel caso non esista alcun node per l'istanza DB/2 desiderata, occorre crearne una con il comando seguente:

CATALOG TCPIP NODE <nodename> REMOTE <dnsname> SERVER <port>

Come <nodename> è consigliabile usare la prima parte del nome DNS del server, designato anche come Hostname. Come <dnsname> è da indicare il nome completo DNS del server. <port> è il numero della porta sulla quale il listener DB/2 si aspetta delle interrogazioni (standard: 50000). Eseguendo ora di nuovo il comando „LIST NODE DIRECTORY“, deve apparire il nuovo <nodename>.

b) ATTACH TO <nodename> USER <siarduser>

Il <nodename> è qui il nome del nodo che indica l'istanza DB/2-Instanz su questo client. Come <siarduser> è da indicare l'UserId dell'utente del database su questa istanza. Eseguito il comando, il sistema chiederà la password per questo utente.

c) LIST DATABASE DIRECTORY

Con questo comando vengono visualizzati tutti i database sull'istanza DB/2, con la quale ci si è connessi.

d) CONNECT TO <database> USER <siarduser>

Come <siarduser> è da indicare l'UserId dell'utente di questo database. Eseguito il comando, il sistema chiederà la password per quest'utente. Tutti gli altri comandi (SQL Statements) si riferiscono a questo database.

Fatto ciò, si può avviare il download con *SIARD Suite* con <dnsname>:<port>/<database> e l'utente <siarduser>.

9 SiardToDb

Il programma *SiardToDb* è un'applicazione a righe di comando, che carica un archivio SIARD in un'istanza di banca dati a scopi di ricerca.

9.1 Avvio

```
java -cp <siardpath>/bin/SiardEdit.jar ch.admin.bar.siard.SiardToDb [-h] |
-c <config file>] -d <data source> -n <database name> -u <database user>
-p <database password> -s <siard file> [-q <query timeout>] [<schema> <mappedschema>]*
```

Sotto *<siardpath>* occorre indicare la cartella nella quale SIARD è stato installato. Nella sottocartella *bin* della cartella d'installazione, si trova il file *SiardEdit.jar* con la classe *ch.admin.bar.siard.SiardToDb*, il cui *main()* viene qui avviato con il comando *java* (sotto Windows è meglio utilizzare *javaw*).

Se c'è l'opzione *-h (help)* sulla riga di comando, viene visualizzata la sintassi d'avvio.

Dato che l'avvio contiene molti parametri, lo si può semplificare creando un file di configurazione nel quale questi vengono memorizzati, ad eccezione della password. Se questo file esiste, l'avvio si riduce a:

```
java -cp <siardpath>/SiardEdit.jar ch.admin.bar.siard.SiardToDb -c <config file>
-p <database password>
```

Se sono stati indicati tutti i parametri e in più un file di configurazione, i parametri nel file di configurazione vengono completati da quelli sulla riga di comando. Ciò semplifica la ripetizione dell'upload di SIARD.

9.2 Parametri d'avvio

<i>parametro</i>	<i>significato</i>
<data source>	ODBC, Oracle, "SQL Server", MySQL o "DB/2"
<database name>	Nome della banca dati, composta diversamente secondo la connessione: ODBC: Datasourcename (DSN) configurato della connessione ODBC Oracle: p.es. dbhost.enternet.ch:1521:SIARD1 SQL Server: p.es. dbhost.enternet.ch:1433/SIARD MySQL: p.es. dbhost.enternet.ch:3306/SIARD DB/2: p.es. dbhost.enternet.ch:50000/SIARD
<database user>	UserId per la sessione di upload
<database password>	Password per la sessione di upload
<siard file>	Nome di file dell'archivio SIARD da caricare
<query timeout>	Timeout per interrogazioni SQL, indicato in secondi (o illimitato)
<schema>	Nome dello schema nell'archivio SIARD (da indicare quante volte si vuole insieme al seguente <mappedschema>)
<mappedschema>	Nome dello schema da utilizzare per l'upload.

<i>parametro</i>	<i>significato</i>
	Per tutti gli schemi, ai quali non è stato assegnato nessun altro nome, viene utilizzato il proprio nome.

9.3 Osservazioni

Dato che le banche dati più vecchie non sono conformi a SQL:1999, non è possibile purtroppo evitare un notevole lavoro di configurazione manuale in preparazione all'upload. Sotto MS Access non ci sono oggetti di schema, mentre sotto SQL Server 2000 e Oracle 8 e 9 utente e oggetti di schema non sono separati.

Durante il caricamento vengono create solo delle tabelle, mentre altri oggetti di banca dati non sono generati. Inoltre il caricamento può avvenire eventualmente con delle riduzioni. In MS Access tutte le tabelle arrivano nella stessa MDB. In Oracle tutti i nomi più lunghi di 30 caratteri vengono abbreviati. Per evitare collisioni, i nomi delle tabelle e delle colonne vengono allungati con un numero (per esempio una tabella „un nome troppo lungo per Oracle“ diventa „un nome troppo lungo per Or10“).

Nei casi in cui la precisione massima e il numero massimo di posti decimali (per esempio in MS Access) sono inferiori al necessario, i valori sono caricati con precisione inferiore. SIARD si adatta per quanto possibile al sistema di banca dati d'arrivo. Consultando i metadati della banca dati con *SiardEdit*, allora è possibile assegnare correttamente denominazioni e valori.

La modifica di TIMES e TIMESTAMPS nella banca dati dipende dal fuso orario locale. Se per esempio si memorizza l'orario 15:30 a Zurigo, nel file XML è memorizzato l'orario UTC 14:30 (in inverno!). Se gli orari della banca dati devono essere interpretati invariabilmente come orari UTC, occorre avviare *SiardFromDb* con l'opzione

```
-Duser.timezone=GMT
```

9.4 Preparazione della banca dati

Secondo il tipo, la banca dati ha bisogno di una diversa preparazione per l'upload.

9.4.1 Preparazione MS Access

Per caricare un file SIARD in una banca dati MS Access, bisogna procedere come segue:

1. creare un file Mdb o accdb vuoto con MS Access.
2. creare una connessione ODBC per l'upload-MDB vuota appena creata (vedi anche la voce ODBC nella guida di Windows sotto *Start / Guida di Windows* per consultare le istruzioni di come creare una connessione ODBC sotto *Start / pannello di controllo / strumenti di amministrazione / origine dati (ODBC)*), e nelle proprietà avanzate (sotto *avanzate ...*) e impostare *ExtendedAnsiSQL=1*. La connessione ODBC può essere impostata da un non-amministratore come *utente DSN*.

Questioni di architettura (32-bit / 64-bit)

Se *SIARD Suite* è stata avviata con un Runtime JAVA a 32 bit, si può accedere solo a un driver ODBC a 32 bit. Se MS Access è installato su una piattaforma Windows a 64 bit, occorre impostare solo accessi ODBC a 64 bit con l'amministratore standard ODBC (*C:\Windows\system32\odbcad32.exe*). Questi sono quindi accessibili a SIARD Suite, soltanto se è stata avviata con un JAVA Runtime a 64 bit. Volendo allestire un accesso ODBC a 32 bit, occorre usare l'amministratore ODBC a 32 bit (*C:\Windows\SysWOW64\odbc-*

SIARD Suite

ad32.exe). Questi accessi ODBC a 32 bit sono accessibili per SIARD, soltanto se è stata avviata con un JAVA Runtime a 32 bit.

I file batch *siardedit.cmd*, *siardfromdb.cmd* e *siardtodb.cmd* (e lo Shortcut impostato sul desktop dall'installazione) utilizzano JAVA Runtime a 64 bit, se questo è stato installato. Se allo stesso tempo è disponibile un JAVA Runtime a 32 bit, si può avviare manualmente *SIARD Suite* con quest'ultimo, richiamando i file batch con il parametro facoltativo */32*. Se sorgono problemi caricando un file MS Access con ODBC a 64 bit, conviene provare con ODBC a 32 bit. Attualmente il driver ODBC a 64 bit ha ancora troppi bugs.

9.4.2 Preparazione Oracle

Per caricare un file SIARD in una banca dati Oracle, bisogna procedere come segue:

1. creare come amministratore di sistema per ogni schema nel file SIARD un utente con lo stesso nome (usare solo maiuscole, poiché i nome d'utente sotto Oracle scritti in minuscolo sono mal supportati) con il comando CREATE USER.
2. assegnare ad un utente dell'archivio (p.es. a uno degli utenti dello schema) i privilegi di sistema CONNECT e RESOURCE.
3. assegnare all'utente dell'archivio tutti i privilegi di creazione e di scrittura su tutti gli schemi.

Il file SIARD contiene solo in casi eccezionali più di uno schema. Nella maggior parte dei casi le istruzioni di cui sopra si riducono alle righe seguenti:

```
CREATE USER <SCHEMA> IDENTIFIED BY <PASSWORD>
GRANT CONNECT, RESOURCE TO <SCHEMA>
```

Durante l'upload è possibile che lunghe stringhe di caratteri (p.es. VARCHAR(4000) su banche dati UTF-16, che accettano al massimo VARCHAR(2000)) vengano trasformate in CLOBs. Inoltre il parametro di inizializzazione NLS_LENGTH_SEMANTICS per l'upload viene definito come CHAR (standard: BYTE).

9.4.3 Preparazione SQL Server

Per caricare un file SIARD in una banca dati SQL Server, bisogna procedere come segue:

1. creare come amministratore di sistema una banca dati vuota per l'upload (in Microsoft SQL Server Management Studio tramite il comando *New Database ...* sotto *Databases* oppure con l'ausilio del comando CREATE DATABASE).
2. creare per ogni schema del file SIARD un *login* con lo stesso nome. A quest'utente di schema non si deve assegnare alcun privilegio di sistema.
3. creare per ogni schema della banca dati upload un utente assegnato al *login*.
4. la banca dati upload assegna ad ogni utente di schema i seguenti privilegi di sistema:
 - a. Alter any role
 - b. Alter any schema
 - c. Alter any user
 - d. Alter
 - e. Connect
 - f. Create function
 - g. Create procedure

SIARD Suite

- h. Create role
- i. Create rule
- j. Create schema
- k. Create table
- l. Create type
- m. Create view
- n. Delete
- o. Execute
- p. Insert
- q. References
- r. Select
- s. Update
- t. View database state
- u. View definition

Sotto SQL Server 2000 gli utenti e gli schemi erano identici, mentre sotto SQL Server 2005/2010/2012 si esegue per ogni schema il comando

```
CREATE SCHEMA <SCHEMA> AUTHORIZATION <SCHEMAUSER>
```

In questo caso è sufficiente assegnare all'utente solo i privilegi create per il database.

9.4.4 Preparazione MySQL

Per caricare un file SIARD in una banca dati MySQL Server, bisogna procedere come segue:

1. creare come amministratore di sistema per ogni schema nel file SIARD un database (=Schema) con il comando CREATE SCHEMA.
2. assegnare all'utente dell'archivio tutti i privilegi di creazione e di scrittura su tutti gli schemi.
3. Il file SIARD contiene solo in casi eccezionali più di uno schema. Nella maggior parte dei casi le istruzioni di cui sopra si riducono alle righe seguenti:

```
CREATE SCHEMA <SCHEMA>;  
CREATE USER '<USER>' IDENTIFIED BY '<PASSWORD>';  
GRANT ALL ON <SCHEMA>.* TO '<USER>';
```

9.4.5 Preparazione DB/2

Per caricare un file SIARD in una banca dati DB/2, bisogna procedere come segue:

1. Creare un utente archivio sul sistema operativo, sul quale gira il DB/2-Service, (p.es.: *siardusr*).

Attenzione: sul sistema operativo LINUX, l'UserId può essere composto solo da lettere minuscole e può essere lungo al massimo otto caratteri!
2. Creare un nuovo database vuoto come amministratore di sistema, che possa ricevere il contenuto del file SIARD.

SIARD Suite

A questo scopo bisogna avviare la finestra con le righe di comando per DB/2 o il command line processor (clp) e digitare quanto segue:

a) LIST NODE DIRECTORY

Questa directory dovrebbe elencare i node registrati sul client. Nel caso non esista alcun node per l'istanza DB/2 desiderata, occorre crearne una con il comando seguente:

```
CATALOG TCPIP NODE <nodename> REMOTE <dnsname> SERVER <port>
```

Come <nodename> è consigliabile usare la prima parte del nome DNS del server, designato anche come Hostname. Come <dnsname> è da indicare il nome completo DNS del server. <port> è il numero della porta sulla quale il listener DB/2 si aspetta delle interrogazioni (standard: 50000). Eseguendo ora di nuovo il comando „LIST NODE DIRECTORY“, deve apparire il nuovo <nodename> .

b) ATTACH TO <nodename> USER <siarduser>

Il <nodename> è qui il nome del nodo che indica l'istanza DB/2-Instanz su questo client. Come <siarduser> è da indicare l'UserId dell'utente del database su questa istanza. Eseguito il comando, il sistema chiederà la password per quest'utente.

c) LIST DATABASE DIRECTORY

Con questo comando vengono visualizzati tutti i database sull'istanza DB/2, con la quale ci si è connessi.

d) CREATE DATABASE <newdatabase> USING CODESET UTF-8 TERRITORY DE

Il nome <newdatabase> per il nuovo database sull'istanza DB/2 può essere scelto a piacimento, a condizione che non esista già e che non sia più lungo di 8 caratteri. Minuscole e maiuscole vengono ignorate.

L'utilizzo del Codeset UTF-8 è utile, altrimenti eventuali righe di caratteri con diresis non entrano nei campi misurati in lunghezze byte da DB/2. L'indicazione del Paese è obbligatoria, si possono però usare altri codici di paesi ISO (p.es. CH, US, ...).

e) CONNECT TO <newdatabase> USER <dbadm>

Come <dbadm> è da indicare l'UserId di un amministratore di database di questa istanza. Eseguito il comando, il sistema chiederà la password per quest'amministratore di database. Tutti gli altri comandi (SQL Statements) si riferiscono a questo database.

f) CREATE SCHEMA <schema> AUTHORIZATION <siarduser>

Per ogni schema del file SIARD. Qui è da indicare come <siarduser> l'UserId creata come utente nel sistema operativo dell'istanza DB/2.

g) Per testare ci si connette come <siarduser> con il nuovo database:

```
CONNECT TO <newdatabase> USER <siarduser>
```

il sistema chiederà la password del <siarduser>.

Queste azioni devono essere eseguite da un amministratore del server DB/2. Per tutto il resto occorre solo l'utente del database con la sua password.

10 Sistemi di banche dati supportati

Nella *SIARD Suite* sono supportati i database „MS Access“, „Oracle“, „SQL Server“, „MySQL“ e „DB/2“. Per l'accesso a MS Access viene utilizzato il ponte JDBC/ODBC di SUN, mentre per gli altri database sono utilizzati i driver JDBC specifici per banche dati.

A causa dell'inadeguatezza dei driver JDBC e ODBC è sconsigliato accedere ad altri database, oltre MS Access, tramite ODBC.

Poiché i diversi sistemi database supportano tipi di dati leggermente differenti, la standardizzazione dei diversi tipi per ogni sistema database adottata da SIARD Suite, è descritta in un documento separato.

11 File di registro

Se sorgono problemi durante l'utilizzo della *SIARD Suite*, è consigliabile salvare un file di registro, che accompagni un messaggio d'errore alla manutenzione. Per protocollare un file di registro, si proceda come segue:

Nella cartella *etc* dell'installazione SIARD si trova il file *logging.properties*. In questo file è deviato il file di registro con livello INFO sulla console. Per deviare il registro in un file, occorre cancellare il carattere di commento # davanti a

```
handlers= java.util.logging.FileHandler, java.util.logging.ConsoleHandler.
```

inoltre nella riga

```
java.util.logging.FileHandler.pattern =logs/siard%u.log
```

si sostituisce il nome del file dopo i segni di uguale con il percorso completo del file di registro desiderato. Occorre tener presente che i backslashes (\) sono da raddoppiare. Sotto Windows si possono utilizzare ovunque gli slashes (/), p.es.

```
java.util.logging.FileHandler.pattern = C:/Projects/Bundesarchiv/SIARD/logs/siard%u.log.
```

In questo file è possibile anche definire il livello di registrazione (.level). I valori consentiti sono: NONE, SEVERE, WARNING, INFO, CONFIG, FINE, FINER, FINEST, ALL. ALL crea il più gran numero di registrazioni, per non sviluppatori incomprensibili, mentre SEVERE crea informazioni solo su errori di sistema. Una buona impostazione per il lavoro quotidiano è INFO.

12 Limitazioni

Il formato SIARD è sottoposto alle seguenti limitazioni:

1. La dimensione del file non può superare i 18'446'744'073'709'551'615 bytes (ca. 18 ExaBytes, limitazione ZIP64).
2. Il numero dei file (tabelle e LOB) non può superare i 4'294'967'295 (ca. 4 miliardi, limitazione ZIP64).
3. Il totale dei metafile del database deve poter essere contenuto nella memoria JAVA.

Le prime due limitazioni non dovrebbero veramente essere tali, poiché i database reali non raggiungeranno queste dimensioni ancora per molto tempo.

La terza limitazione invece, in circostanze sfavorevoli come poca memoria disponibile e complessità dei metadati del database, può senz'altro verificarsi. Questa limitazione può essere aggirata, lanciando le applicazioni SIARD su un sistema che abbia sufficiente memoria (p.es. 4 GB) e che ingrandisca eventualmente l'heap JAVA con il parametro della riga di comando JAVA „-Xmx2000m“ o simili sul comando JAVA.