



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI

Schweizerisches Bundesarchiv BAR

Abteilung Informationsüberlieferung

Dienst Sicherung und Archivierungslösungen

Siard Suite Manuale

Version 2.2.136-SNAPSHOT

Contenuti

1. Introduzione	1
2. Contesto	1
3. Requisiti di sistema e installazione	1
3.1. Installazione ed esecuzione dell'applicazione.....	2
4. Utilizzo della suite SIARD	2
4.1. GUI della suite SIARD.....	2
4.2. SIARD Suite CLI	5
4.3. SiardFromDb	5
4.4. SiardToDb	7
5. Suggerimenti e trucchi	9
5.1. Preparazione di un database per il download.....	9
5.2. Preparazione di un database per il caricamento	9
5.3. Tempo, timestamp e fusi orari.....	9
6. Restrizioni	10

1. Introduzione

Lo standard SIARD (Software-Independent Archival of Relational Databases) definisce un formato per l'archiviazione a lungo termine dei contenuti dei database relazionali. L'applicazione "SIARD Suite" fornisce un'interfaccia grafica per l'archiviazione, il ripristino, la ricerca e l'esportazione di archivi SIARD.

2. Contesto

In collaborazione con i progetti europei E-ARK e KOST, la versione 2.1 del formato SIARD è stata specificata tra il 2015 e il 2018. Come la versione 1.0, il formato SIARD 2.1 è stato omologato come standard eCH-0165 dall'associazione "eCH E-Government Standards". Nel 2021, il Consiglio DILCIS ha sviluppato e approvato il formato SIARD 2.2 esteso.

SIARD Suite è l'implementazione di riferimento per l'archiviazione di database relazionali nel formato standardizzato SIARD 2.2. Inoltre, SIARD Suite può anche leggere archivi SIARD nei formati SIARD 2.1 e SIARD 1.0 e convertirli in SIARD 2.2.

3. Requisiti di sistema e installazione

L'applicazione SIARD Suite è un'applicazione Java FX. Come tale, può essere eseguita su qualsiasi sistema su cui sia installato un Java Runtime Environment 8 (JRE) con l'estensione JavaFX.

Inoltre, sono disponibili distribuzioni con JRE incluso e installatori nativi per Windows, Mac e Linux, in modo che non sia necessario installare JRE sul sistema.

Le singole distribuzioni sono elencate di seguito. Tutte le distribuzioni possono essere scaricate qui: <https://github.com/sfa-siard/siard-suite/releases>

Nome della distribuzione	Descrizione
SIARD-Suite-v2.2.133	Distribuzione senza runtime, richiede l'installazione di JRE con JavaFX sul sistema
SIARD-Suite-v2.2.133-Linux	Distribuzione con runtime incluso per Linux
SIARD-Suite-v2.2.133-Windows	Distribuzione con runtime incluso per Windows
SIARD-Suite-v2.2.133-macOS	Distribuzione con runtime incluso per MacOS
SIARD-Suite-v2.2.133-Linux-deb	Installatore nativo per Linux
SIARD-Suite-v2.2.133-win-exe	Installatore nativo per Windows (exe)
SIARD-Suite-v2.2.133-win-msi	Installatore nativo per Windows (msi)
SIARD-Suite-v2.2.133-macOS-dmg	

Nota: Se si utilizza il programma di installazione nativo, la Siard CLI (Command Line Interface) non è disponibile. Per poter utilizzare la CLI, è necessario utilizzare una distribuzione con o senza JRE inclusa.

3.1. Installazione ed esecuzione dell'applicazione

Se si utilizzano gli installatori nativi, l'installazione avviene semplicemente tramite il sistema. Se necessario, si utilizzano i diritti di amministratore. I programmi di installazione creano anche i collegamenti corrispondenti per avviare l'applicazione.

Le distribuzioni con e senza JRE incluso sono fornite come archivio ZIP. Questo può essere decompresso in qualsiasi posizione. L'applicazione può quindi essere avviata tramite uno degli script di avvio:

Nome dello script	descrizione
siard-suite (.bat)	Avvio dell'interfaccia grafica per sistemi Unix o Windows
siard-from-db (.bat)	Archiviazione di un database (CLI)
siard-to-db (.bat)	Caricamento di un archivio SIARD in un database (CLI)

4. Utilizzo della suite SIARD

La suite SIARD è costituita dai seguenti componenti:

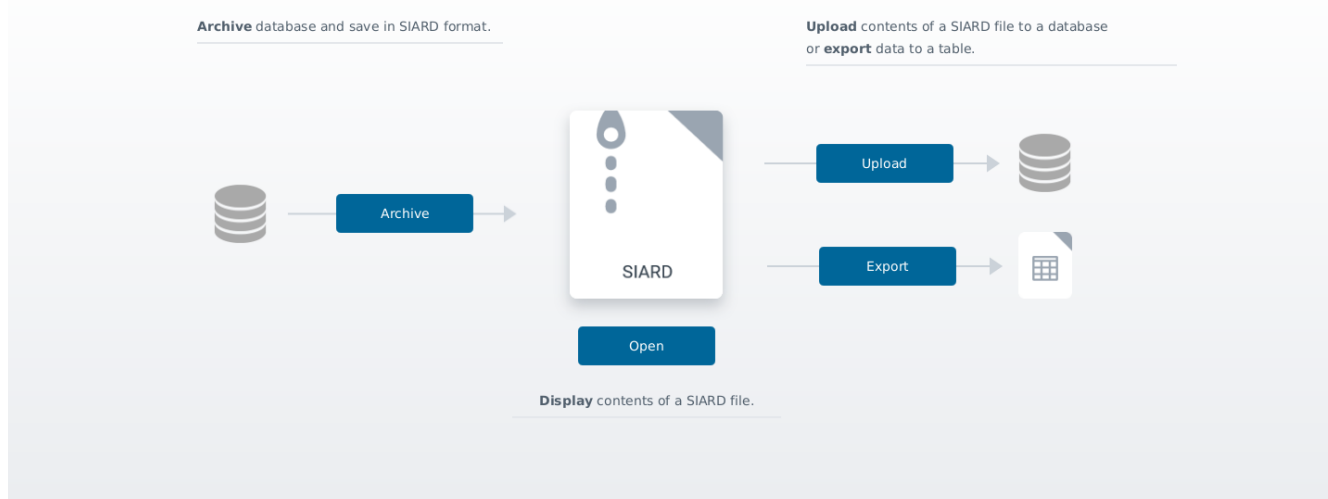
- Interfaccia grafica utente (GUI)
- SiardFromDb (CLI)
- SiardToDB (CLI)

Lo scopo e l'uso dei componenti sono spiegati di seguito.

4.1. GUI della suite SIARD

L'interfaccia grafica è l'applicazione centrale per la creazione di archivi SIARD e la loro successiva visualizzazione senza modificare i dati primari dell'archivio. Per analisi complesse dei dati archiviati, l'archivio deve essere caricato nuovamente in un database.

La schermata di avvio costituisce il punto di accesso all'applicazione. Qui è possibile avviare le quattro funzionalità principali dell'applicazione:



- Archivio database
- Aprire l'archivio SIARD
- Caricare l'archivio SIARD in un database
- Esportazione di dati da un archivio SIARD

L'applicazione guida l'utente attraverso i rispettivi processi.

4.1.1. Archiviazione di database con dati binari

Lo standard SIARD Format 2.2 prevede che gli oggetti di grandi dimensioni (LOB) di un database possano essere archiviati nel file system esterno.

La posizione di memorizzazione dei dati può essere registrata durante l'archiviazione nella fase "metadati".

Metadati

Modifica i metadati per il tuo archivio

Cosa verrà archiviato?

Nome database

testdb

Descrizione del database (opzionale)

Ufficio di consegna: proprietario dei dati prima dell'archiviazione

Ora di origine

Chi ha archiviato?

Il tuo nome (opzionale)

I tuoi dati di contatto (es. e-mail, numero di telefono) ⓘ

Cartella per esportare oggetti di grandi dimensioni (opzionale)

☐ Esporta viste come tabelle

Ritorno

Aborto

Continua

Salva solo i metadati

Il percorso può essere specificato in modo assoluto come URI di file (**file:/**) o relativo, ad esempio **../lobs/** - in ogni caso il percorso deve terminare con una barra (/). Si consiglia di specificare percorsi relativi, in modo che l'archivio possa essere spostato insieme ai dati LOB esterni; ciò non è possibile se si utilizzano percorsi assoluti. La struttura all'interno della cartella specificata viene generata automaticamente e segue la struttura del database: **schema/tabella/colonna**. SIARD Suite rileva automaticamente il tipo MIME e le estensioni dei file vengono impostate automaticamente in modo corretto. Questo permette anche di aprire i dati binari direttamente da SIARD Suite nell'applicazione corretta installata sul sistema.

Impostazioni di ==

L'applicazione può essere configurata alla voce di menu "Opzioni".

Options

X

Login Timeout (s)

Query Timeout (s)

Cancel

Save

È possibile effettuare le seguenti impostazioni:

Opzione	Descrizione
Timeout di accesso (s)	Tempo massimo di attesa per una connessione al database (in secondi). 0 per illimitato
Timeout di interrogazione (s)	Tempo massimo di attesa per un'interrogazione al database (0 per illimitato).

4.2. SIARD Suite CLI

Oltre all'interfaccia grafica, è possibile archiviare i database o rileggere un archivio in un database tramite la riga di comando.

A questo scopo, nella distribuzione ZIP sono disponibili gli script `siard-from-db` e `siard-to-db`.

4.3. SiardFromDb

Con `siard-from-db` è possibile archiviare un database senza utilizzare la GUI.

4.3.1. Chiamata

In un terminale, passare alla directory in cui è stato scompattato l'archivio ZIP, aprire la directory `bin` lì ed eseguirla con

```
siard-from-db [-h] | [-o] [-v] [-l=<login timeout>] [-q=<query timeout>] [-i=<importare meta dati>] [-x=<cartella lob esterna>] [-m=<tipo di file>] -j=<URL JDBC> -u=<utente del database> -p=<password del database> -s=<file siard> -e=<esportare meta dati>".
```

4.3.2. Argomenti

Argomento	descrizione
-o	sovrascrive i file di output, se disponibili
-v	Archivia le viste come tabelle
<login timeout>	Timeout in secondi per il login (0 per illimitato)
<query timeout>	Timeout in secondi per la query (0 per illimitato)
<import meta data>	Nome del file XML dei meta dati da usare come modello.
<cartella LOB esterna>	Cartella in cui memorizzare esternamente i dati della colonna LOB (large object) più grande del database (i contenuti con lo stesso nome verranno sovrascritti). <mime type>
Tipo MIME dei dati della colonna LOB più grande del database (influenza la scelta dell'estensione del nome dei LOB memorizzati esternamente)	<JDBC URL>
URL JDBC del database da scaricare. Esempi: MS Access: jdbc:access:<path>/<filename>.mdb DB/2: jdbc:<host>:50000/<nome-database> MySQL: jdbc:mysql://<host>:3306/<nome-del-db> Oracle: jdbc:oracle:thin:@<host>:1521:orcl postgres: jdbc:postgresql://<host>:5432/<nome-del-db> Server SQL: jdbc:sqlserver://<host>/<db-name>:1433	<utente del database>
utente del database	<password del database>
password del database	<file siard>
nome del file .siard da creare	<export meta data>

4.3.3. Note

È necessario impostare almeno uno dei due parametri **siard-file** o **export meta data** (è anche possibile utilizzare entrambi i parametri).

Nota: Durante il processo di archiviazione, il contenuto del database non deve essere modificato: se

necessario, utilizzare un'istantanea del database.

L'opzione Archivia le viste come tabelle di solito porta a una memorizzazione ridondante degli stessi dati ed è quindi sconsigliata. È utile se l'utente del database utilizzato per l'archiviazione può visualizzare il contenuto delle viste ma non quello delle tabelle di base.

Il processo di archiviazione o riesce completamente o fallisce.

Per i database di grandi dimensioni, si consiglia di scaricare prima l'XML dei metadati. Questo contiene tutti i metadati e le dimensioni delle tabelle e aiuta quindi a stimare il tempo di esecuzione necessario per il download. Inoltre, per le tabelle di grandi dimensioni, si dovrebbe usare l'opzione -q 0, poiché non è possibile stimare quanti secondi impiegherà una query sulle dimensioni.

4.3.4. Utente del database

In genere si sconsiglia di utilizzare l'amministratore del database (DBA, root, dbo, ...) con tutti i diritti per scaricare un archivio SIARD. L'ambito dei dati memorizzati nell'archivio SIARD è definito dagli oggetti a cui l'utente dell'archivio ha accesso in lettura. Il DBA globale può normalmente leggere tutti i database del sistema e anche molte tabelle di sistema che non devono essere archiviate. Per questo motivo, è importante utilizzare un utente di archivio adatto per il download, che potrebbe dover essere creato ex novo a questo scopo.

4.4. SiardToDb

SiardToDb è un programma a riga di comando che carica un archivio SIARD in un database per scopi di ricerca.

4.4.1. Chiamata

```
siard-to-db [-h] [-o] [-l=<login timeout>] [-q=<query timeout>]  
-s=<file siard> -j=<JDBC-URL> -u=<database-user> -p=<database-password>  
[<schema> <mappedschema>]
```

4.4.2. Argomenti

argomento	descrizione
-o	Sovrascrive i tipi e/o le tabelle con lo stesso nome nel database, se presenti.
<login timeout>	Timeout in secondi per il login (0 per illimitato)
<query timeout>	Timeout in secondi per una richiesta (query) (0 per illimitato)
<file siard>	Nome del file .siard da caricare

<JDBC URL>	URL JDBC del database di destinazione Esempi: MS Access: *jdbc:access:<path>/<nome-file>.mdb *DB/2: jdbc:<host>:50000/<nome-database> MySQL: jdbc:mysql://dbserver.enterag.ch:3306/<nome-del-db> Oracle: jdbc:oracle:thin:@<host>:1521:orcl postgres: jdbc:postgres://<host>:5432/<nome-del-db> Server SQL: jdbc:sqlserver://<host><nome-del-db>:1433
<utente del database>	utente del database
<password del database>	password del database
<schema>	nome dello schema nell'archivio SIARD
<mappedschema>	nome dello schema da utilizzare nel database

4.4.3. Note

Poiché i vecchi database non sono conformi a SQL:2008, è inevitabile che sia necessaria una preparazione manuale per il caricamento. In MS Access non esistono oggetti schema. In Oracle, utenti e schemi sono inseparabili. In MySQL, gli oggetti schema e i database sono la stessa cosa. Pertanto, tutti gli schemi devono essere creati manualmente prima del caricamento. Inoltre, l'utente del database deve avere il diritto di creare tabelle e tipi in questi schemi. Poiché questo non è sempre facile, gli schemi del database sono assegnati ai nomi degli schemi in base all'elenco di mappatura sulla riga di comando.

Il caricamento crea solo tabelle e tipi e cerca di attivare chiavi primarie e straniere. Non vengono creati altri oggetti nel database di destinazione. Se le chiavi non possono essere attivate, viene emesso un avviso, ma il caricamento viene comunque considerato riuscito. Le query SQL SELECT sul database possono essere eseguite anche senza queste chiavi.

Inoltre, potrebbero esserci delle restrizioni al momento del caricamento. Con MS Access, tutte le tabelle finiscono nello stesso MDB/ACCDB. Con Oracle, tutti i nomi superiori a 30 caratteri vengono accorciati. Per evitare collisioni, ai nomi delle tabelle e delle colonne viene assegnato un contatore. (Ad esempio, "Un nome di tabella troppo lungo per Oracle" diventa "Un nome di tabella Table01").

Se la precisione massima e il numero massimo di cifre decimali (ad esempio in MS Access) sono inferiori a quelli richiesti, i valori vengono caricati con una precisione inferiore. SIARD si aiuta il più possibile nel sistema di database di destinazione. Se i metadati del database vengono consultati con SiardGui, è possibile assegnare le etichette e i valori corretti.

5. Suggerimenti e trucchi

5.1. Preparazione di un database per il download

SIARD Suite scarica tutti i dati visibili all'utente del database utilizzato per la connessione. La scelta di un utente di database adatto determina quindi la portata del database archiviato. Spesso è disponibile un "utente tecnico del database" di un'applicazione di database con cui viene amministrato il database. Questo utente di solito ha esattamente i diritti di accesso necessari per l'archiviazione del database.

Tuttavia, se non è disponibile un utente di database di questo tipo, l'amministratore del database (DBA, dbo, root, SYSTEM, sa, ...) non dovrebbe essere utilizzato per il download. Questo perché l'amministratore del database ha accesso in lettura a molte tabelle di sistema e a tabelle di altri database che non devono essere archiviati. Si consiglia invece di creare un utente del database per l'archiviazione. A questo utente di archiviazione deve essere concesso l'accesso in lettura a tutti gli schemi, le tabelle, le viste e i tipi appartenenti al database da archiviare. Le modalità di creazione di un utente e di assegnazione dei privilegi necessari sono riportate nella documentazione del DBMS.

Una volta determinato o creato l'utente adatto, il download del database può essere eseguito con i dati di accesso dell'utente.

5.2. Preparazione di un database per il caricamento

Per caricare un database su un'istanza di un DBMS con SIARD Suite, nel database di destinazione devono essere disponibili schemi di database adeguati. Questi schemi possono essere mappati agli schemi dell'archivio SIARD nella mappatura degli schemi della finestra di dialogo di upload (o nella mappatura degli schemi della riga di comando SiardToDb).

Idealmente, gli schemi di destinazione sono vuoti. Tuttavia, a seconda delle impostazioni di sicurezza del DBMS, non è sempre possibile creare o selezionare schemi a piacere. Se gli schemi di destinazione non sono vuoti, i dati archiviati verranno caricati solo se non si verificano collisioni di nomi per i tipi e le tabelle o se l'opzione Sovrascrivi è stata esplicitamente selezionata.

L'utente del database i cui dati di accesso vengono utilizzati per il caricamento deve avere il diritto di creare tipi e tabelle negli schemi di destinazione e di inserire dati in queste tabelle. La documentazione del DBMS contiene informazioni su come creare gli schemi e su come concedere all'utente del database i diritti necessari. Se possibile, è più facile utilizzare i dati di accesso dell'amministratore del database. Quest'ultimo, infatti, dispone solitamente di tutti i privilegi necessari. In questo caso, tuttavia, l'opzione Sovrascrivi non deve essere selezionata. In caso contrario, il rischio di sovrascrivere tabelle o tipi importanti del sistema è troppo elevato.

5.3. Tempo, timestamp e fusi orari

La conversione di TIME e TIMESTAMP nel database dipende dal fuso orario locale. L'ora UTC 14:30 nell'archivio SIARD viene caricata nel database come ora locale 15:30 quando viene caricata a Zurigo - in inverno. Per sopprimere questa conversione, SiardToDb deve essere avviato con l'opzione `Duser.timezone=GMT`, che fa sì che SIARD interpreti tutti gli orari del database come orari

6. Restrizioni

Il formato SIARD è soggetto alle seguenti restrizioni

- La dimensione di un file SIARD non deve superare i 18'446'744'073'709'551'615 byte (circa 18 ExaByte) (limitazione ZIP64).
- Il numero di file (tabelle e LOB) non deve superare i 4.294.967.295 (circa 4 miliardi) (limitazione ZIP64).

La suite SIARD è inoltre limitata da

- L'intero metadato del database deve essere contenuto nella memoria JAVA (heap). Questa limitazione può verificarsi in una costellazione sfavorevole di memoria principale disponibile e complessità dei metadati del database. Può essere aggirata avviando le applicazioni SIARD su una macchina con sufficiente memoria principale e, se necessario, ampliando l'heap JAVA con il parametro della riga di comando JAVA -Xmx2000m o un parametro simile sulla riga di comando JAVA.