



Hartwig Thomas, 4 décembre 2013

Version du manuel 1.20

SIARD Suite

Manuel



Edité par:

Archives Fédérales Suisses
Archivstrasse 24
3003 Bern
Suisse

Résumé

Ce document est un manuel technique d'utilisation pour l'application *SIARD Suite* (Software-Independent Archival of Relational Databases) des Archives Fédérales Suisses.

Il décrit

- Les conditions techniques pour son utilisation
- L'installation
- L'exécution

de la *SIARD Suite* et ses composants.

1 Table des matières

1	TABLE DES MATIÈRES	3
2	INTRODUCTION	4
2.1	DROITS D’AUTEURS	4
3	CONCEPT SIARD	5
4	CONDITIONS D'UTILISATION DE LA <i>SIARD SUITE</i>	6
4.1	JAVA	6
4.2	BASES DE DONNÉES	6
5	INSTALLATION	8
6	COMPOSANTS SIARD	9
7	SIARDEDIT	10
7.1	PREMIER DÉMARRAGE	10
7.2	FENÊTRE PRINCIPALE	11
7.3	MENUS	14
8	SIARDFROMDB	25
8.1	COMMANDE	25
8.2	PARAMÈTRES DE COMMANDE	25
8.3	REMARQUES	26
8.4	PRÉPARATION DE LA BASE DE DONNÉES	26
9	SIARDTODB	31
9.1	COMMANDE	31
9.2	PARAMÈTRES DE COMMANDE	31
9.3	REMARQUES	32
9.4	PRÉPARATION DE LA BASE DE DONNÉES	32
10	SYSTÈMES DE BASES DE DONNÉES SUPPORTÉS	37
11	PROTOCOLE	38
12	RESTRICTIONS	39

2 Introduction

Dans le cadre du projet ARELDA (de l'allemand ARchivierung ELEktronischer Daten und Akten), les Archives Fédérales Suisses ont développé l'application *SIARD Suite* (Software-Independent Archival of Relational Databases) servant à l'archivage à long terme de contenus de bases de données.

Sur mandat des Archives fédérales, Enter AG a recueilli des expériences pratiques en matière d'archivage de bases de données productives (2005, 2006) et, à la fin de l'année 2007, a développé la *SIARD Suite*. Cette dernière englobe les expériences pratiques faites dans un but d'amélioration du produit.

Le présent document représente le manuel d'utilisation de la *SIARD Suite*.

2.1 Droits d'auteurs

La *SIARD Suite* est un développement des Archives Fédérales Suisses. Tous les droits sont réservés aux Archives Fédérales Suisses.

La *SIARD Suite* se base sur ce composant d'autres éditeurs :

- JavaHelp 2.0.05 (de <http://java.sun.com/products/javahelp>,

License: dans le dossier SIARD sous *doc/JavaHelp-LICENSE.HTML*)

Ce composant est délivré sur le CD SIARD sous forme de fichier JAR afin d'en faciliter l'installation. Les utilisateurs de la *SIARD Suite* sont tenus de respecter les clauses de licences présentes dans le dossier */doc*.

3 Concept SIARD

Comme les Archives Fédérales Suisses sont tenues d'archiver les documents de l'administration fédérale, indépendamment des supports d'informations¹, la question de l'archivage à long terme de bases de données relationnelles doit être résolue².

Dans le cadre du projet ARELDA, on est arrivé à la conclusion que les systèmes actuels de bases de données étaient intégrés de manière trop propriétaire et qu'ainsi, sur cette seule base, aucun archivage à long terme n'était possible.

Afin d'assurer une disponibilité à long terme, il a été décidé de mémoriser les bases de données conformément à SQS:1999³. Les contenus ont été enregistrés dans une collection de fichiers XML. Etant donné que le format d'archivage pour bases de données est basé sur des standards ISO, on lui attribue une compréhension de plus longue durée.

Dans les chapitres suivants, le format d'archivage pour bases de données est nommé format SIARD.

Une exigence importante pour les données archivées en format SIARD est que les contenus des bases de données aient le caractère de "documents". Cela signifie que les contenus des tableaux contiennent les données d'affaires exploitées par la base de données et qui demeurent compréhensibles indépendamment des applications frontend. Avec la *SIARD Suite*, ni le code de programme exécutable, ni les objets ne sont archivés, mais uniquement les données d'affaires dans des tables de bases de données.

La définition du schéma XML d'une base de données archivée en format SIARD est enregistrée dans des fichiers XML conformément à la norme SQL:1999. Par contre, la documentation des contenus des tables et champs ainsi que les données sont enregistrées dans des fichiers XML. Afin que ces derniers ne soient pas remplis démesurément, les données BLOB (Binary Large Objects) et CLOB (Character Large Objects) sont enregistrées dans des fichiers (binaires) séparés qui sont référencés dans les fichiers XML.

Dans la *SIARD Suite*, actuellement seules les fonctions centrales, les „Core Features“ de SQL:1999, sont utilisées. Dans de prochaines versions de la *SIARD Suite*, l'assistance pourra être étendue à d'autres éléments appelés „Packages“ de SQL:1999.

Ce manuel ne s'étend pas plus loin sur la structure du format SIARD, car ceux-ci sont décrits dans un autre document livré avec la *SIARD Suite*. En 2013, le format SIARD a été reconnu comme standard 2CH et peut être téléchargé sous eCH-0165⁴. Dans la version actuelle de SIARD Suite, le format SIARD est encore différent du standard eCH quant à l'URL.

¹ Loi fédérale sur l'archivage (loi sur l'archivage, BGA) du 26 juin 1998 (état au 22 Juillet 2003)

² Archivage de données électroniques numériques et actes de l'administration fédérale dans les Archives Fédérales Suisses (ARELDA), exposé du projet du 1.3.2001

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

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=fr>

4 Conditions d'utilisation de la *SIARD Suite*

L'utilisation de la *SIARD Suite* nécessite l'installation de JAVA. En outre, une infrastructure adéquate est requise pour l'échange de données (chargement ou extraction) avec un système de bases de données.

4.1 JAVA

Les utilisateurs de la *SIARD Suite* doivent installer préalablement JAVA. La condition minimale technique est JRE 1.56.

JAVA peut être obtenu gratuitement sur le site <http://www.java.com/de/> (JRE – JAVA Runtime Environment) ou <http://www.oracle.com/technetwork/java/javase/downloads/index.html> (JDK - JAVA Development Kit). Si l'on n'est pas intéressé à télécharger des bases de données MS Access, on peut, à la place, utiliser aussi OpenJDK (<http://openjdk.java.net/>).

Avant la version 1.5, JAVA n'a malheureusement pas la possibilité centrale de détecter la version JAVA installée. Il est même possible d'avoir plusieurs versions JAVA installées. Ainsi, le client Oracle installe sa propre version JAVA (souvent plus ancienne et comportant des erreurs) et l'insère dans les variables d'environnement PATH (!).

Il est possible de contrôler si JAVA 1.56 ou une version plus récente est disponible sous Windows de la manière suivante: sous *Démarrer / Exécuter*, taper l'instruction *regedit*. Si JAVA 1.5 ou plus récent est installé, la clé *CurrentVersion* dans *HKEY_LOCAL_MACHINE\SOFTWARE\JavaSoft\Java Runtime Environment* a la valeur "1.56" ou plus haute. Si la branche du registre n'existe pas ou qu'une *CurrentVersion* plus ancienne est indiquée, une nouvelle version JAVA doit être installée.

4.1.1 Architecture (32-bit/64-bit)

Dès JAVA 1.6, il est possible d'installer sur un système d'exploitation 64-bit (par ex. Windows7, Ubuntu, MacIntosh) une version 64-bit de JAVA. Ceci est généralement recommandé. Le code JAVA pour *SIARD Suite* ne dépend pas du fait qu'il soit exécuté sur la version 32-bit ou 64-bit du Runtime JAVA. Il y a toutefois une dépendance à l'architecture du système d'exploitation dans les ponts JDBC-ODBC qui est utilisée pour les bases de données MS Access⁵. Si JAVA 64-bit est utilisé, MS Access 64-bit doit aussi être installé (resp. le pilote ACE ODBC). Avec JAVA 64-bit, il n'est pas possible d'accéder aux sources de données ODBC 32-bit. Malheureusement, le pilote ODBC 64-bit de Microsoft comporte encore beaucoup d'erreurs, ce qui rend impossible la conversion de certains fichiers MS Access.

Si JAVA 32-bit et JAVA 64-bit sont installés, *SIARD Suite* démarre par défaut avec le runtime 64-bit. Il peut toutefois être lancé manuellement aussi avec le runtime 32-bit.

4.2 Bases de données

Dans la livraison de base, *SIARD Suite* supporte les systèmes de bases de données suivants:

- Oracle
- SQL Server
- MySQL (ou MariaDB)

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

SIARD Suite

- DB/2
- MS Access

Ultérieurement, d'autres systèmes de bases de données pourront être ajoutés.

La *SIARD Suite* peut accéder aux bases de données au moyen des pilotes JDBC suivants:

sqljdbc4.jar	Seulement pour accès aux bases de données SQL Server, SQL Server 2000 à SQL Server 2012
ojdbc6-fixed.jar	Seulement pour accès aux bases de données Oracle, Oracle 8i à Oracle 11g (corrige des erreurs dans la distribution Oracle, qui empêchent la création de fichiers XSD, dans laquelle seul des packages Oracle étaient conservés.
mysql-connector-java-5.1.18-bin.jar	Seulement pour l'accès aux bases de données MySQL 5.5 ou plus récent.
sunjdbc.jar	Pour l'accès aux bases de données MS Access via Open Database Connectivity de Microsoft ODBC 3.5
db2jcc.jar et db2jcc_license_cu.jar	Pour l'accès aux bases de données DB/2 version 8 ou plus récente.

Ces pilotes sont livrés avec la *SIARD Suite* et installés. Il est très probable qu'ils demeureront compatibles encore plusieurs années avec de futures versions de systèmes de bases de données.

5 Installation

La *SIARD Suite* peut être lancée à tout moment depuis un stick USB, un CD-ROM ou en réseau, pour autant que JAVA 1.56 ou plus récent ait été correctement installé. Ainsi, elle peut fonctionner sous Windows, LINUX, MacIntosh, etc.

Les réglages spécifiques de l'utilisateur sont enregistrés sous `<home>/java/siard suite.properties`. `<home>` désigne ici le dossier de base de l'utilisateur, indépendant de la plateforme (sous Windows XP, il s'agit en principe de `C:\Documents and Settings\<userid>`), sous Windows 7 il s'agit de `C:\Users\<userid>`.

La *SIARD Suite* est livré sans programme d'installation spécifique. Le programme principal *SiardEdit* détecte où la *SIARD Suite* a été installé une fois chez l'utilisateur et de quelle version il s'agit.

Pour installer la *SIARD Suite*, il faut donc démarrer simplement *SiardEdit* – soit avec un script spécifique à la plateforme dans le dossier principal de produit livré (*siardedit.cmd* resp. *siardedit.sh*) ou comme application JAVA en exécutant `java -jar bin/SiardEdit.jar`.

Si *SiardEdit* constate qu'il n'y a encore aucun réglage spécifique à l'utilisateur, il propose l'installation sur le système de l'utilisateur. Le dossier standard imposé pour l'installation est `<home>/applications/siard suite`.

On peut désinstaller à tout moment la *SIARD Suite* sans droit d'administrateur dans *SiardEdit* sous le menu *Options* ou l'installer dans un nouveau dossier.

Même si une ancienne version de la *SIARD Suite* est installée dans le système de l'utilisateur, on peut lancer une nouvelle version depuis un stick USB, un CD-ROM ou le réseau, sans modifier l'ancienne version.

Si *SIARD Suite* est installé sous Windows, *SiardEdit* crée automatiquement un raccourci sur le bureau ainsi qu'un autre sous *Start/Programme* à l'aide du script *desktop.cmd*.



6 Composants SIARD

La *SIARD Suite* comprend les composants suivants:

- SiardEdit
- SiardFromDb
- SiardToDb

Le programme *SiardEdit* possède une interface utilisateur graphique interactif. Il permet le traitement de métadonnées dans une archive SIARD.

Le programme *SiardEdit* est l'outil central avec lequel le traitement de données en format SIARD s'effectue. Les données primaires ne peuvent être modifiées. *SiardEdit* n'est pas adapté pour des recherches complexes. Pour ces dernières, il est recommandé de charger l'archive SIARD dans une base de données et d'effectuer les recherches à l'aide des outils de la base de données.

L'application lignes de commande *SiardFromDb* sert à télécharger une base de données et de l'enregistrer dans un fichier SIARD. La fonctionnalité de cette application est identique aux fonctions disponibles dans *SiardEdit* sous le menu *Nouvelle archive de base de données...* L'utilisation de la version lignes de commandes est plus confortable surtout lors du téléchargement de grosses bases de données. Avec une déviation de *stdout* et *stderr* le téléchargement de longue durée peut être mieux documenté (voir "Utilisation d'opérateurs de déviation de commandes" dans l'aide Windows sous *Démarrer* ou online⁶).

L'application lignes de commande *SiardToDb* sert au chargement d'une base de données depuis un fichier SIARD. La fonctionnalité de cette application est identique avec celle disponible dans *SiardEdit* sous le menu *Charger dans bases de données...* L'utilisation de la version lignes de commandes est plus confortable surtout lors du chargement de grosses bases de données. Avec une déviation de *stdout* et *stderr* le chargement de longue durée peut être mieux documenté (voir "Utilisation d'opérateurs de déviation de commandes" dans l'aide Windows sous *Démarrer* ou online⁷).

⁶ [http://technet.microsoft.com/de-de/library/cc772622\(WS.10\).aspx](http://technet.microsoft.com/de-de/library/cc772622(WS.10).aspx)

⁷ [http://technet.microsoft.com/de-de/library/cc772622\(WS.10\).aspx](http://technet.microsoft.com/de-de/library/cc772622(WS.10).aspx)

7 SiardEdit

Le programme *SiardEdit* possède une interface utilisateur graphique interactif. Il permet le traitement de métadonnées dans une archive SIARD.

SiardEdit permet de:

- Créer une archive SIARD "vide" (sans données primaires) pour se donner une première idée du processus d'archivage,
- De comparer les métadonnées d'une archives SIARD à créer avec les métadonnées d'un autre modèle (vide ou non) d'archive (reprendre toutes les descriptions des éléments qui existent dans une nouvelle archive ou dans un modèle),
- Afficher les métadonnées d'une archive SIARD, de les trier différemment et de les échantillonner, de les modifier ou les compléter, pour autant que cela ne touche pas les données primaires enregistrées dans une archive SIARD,
- Afficher les données primaires d'une archive SIARD, de les trier différemment et de les échantillonner.

Le programme *SiardEdit* est l'outil central avec lequel le traitement de données en format SIARD s'effectue. Les données primaires ne peuvent être modifiées. *SiardEdit* n'est pas adapté pour des recherches complexes. Pour ces dernières, il est recommandé de charger l'archive SIARD dans une base de données et d'effectuer les recherches à l'aide des outils de la base de données.

La conversion des TIMES ou TIMESTAMPS dans la base de données dépend de la zone horaire locale. Lorsqu'on enregistre par exemple l'heure 15.30 à Zurich, l'heure UTC 14.30 (en hiver!) sera enregistrée dans le fichier XML. Si les heures inchangées dans la base de données doivent être interprétées comme heures UTC, il faut démarrer *SiardFromDb* avec l'option

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

7.1 Premier démarrage

La *SIARD Suite* est disponible en fichier ZIP à télécharger. Ce fichier doit tout d'abord être décompressé. Dans le dossier *bin* de la livraison se trouve le fichier *SiarEdit.jar*. Si JAVA est correctement installé, on peut exécuter ce programme sous Windows en effectuant un double clic sur ce fichier. On peut aussi exécuter le script propre à la plateforme *siardedit.cmd* (pour Windows) ou *siardedit.sh* (pour LINUX).

Si cela ne fonctionne pas ou que l'on travaille sous un autre système d'exploitation, on peut exécuter *SiardEdit* également depuis la ligne de commande dans le dossier *bin* de la *SIARD Suite* de la manière suivante:

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

```
java -jar SiardEdit.jar (toutes les plateformes)
```

La condition ici est que le dossier *bin* de l'installation JAVA existe dans le *PATH*. Sinon il faut indiquer le chemin d'accès complet du programme exécutable *java*.

Lorsqu'on exécute pour la première fois *SiardEdit*, le message suivant apparaît:

SIARD Suite



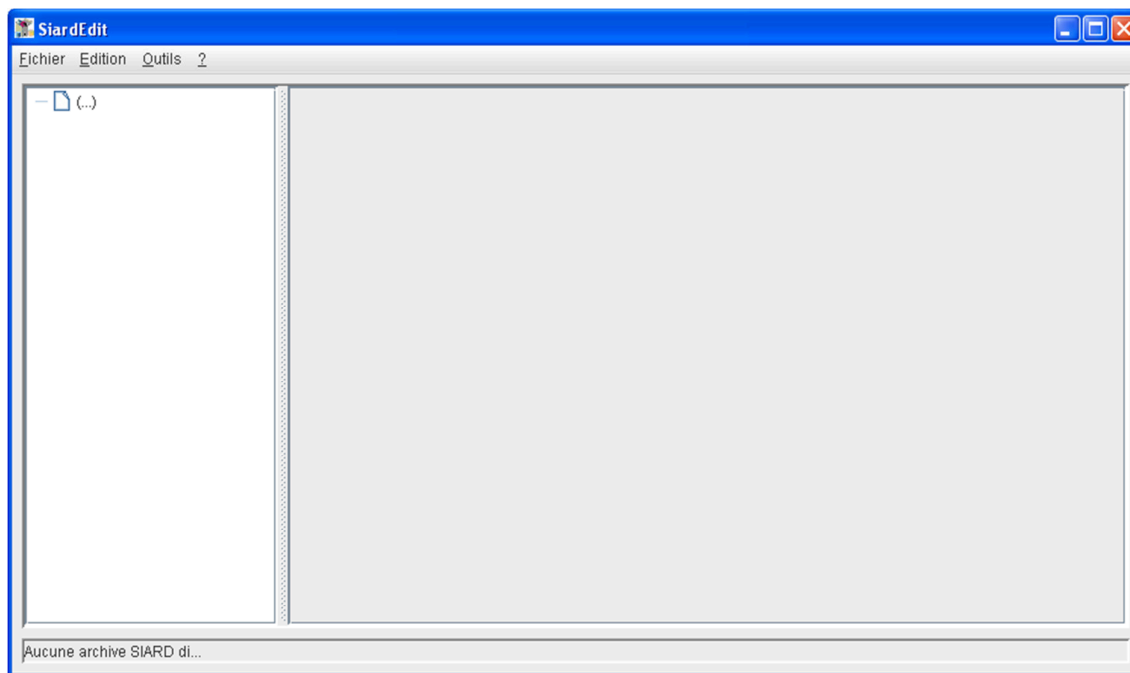
(Comme à cet instant *SiardEdit* ne connaît encore pas la langue d'utilisation, la langue de ce message dépend de celle du système d'exploitation ainsi que de celle choisie lors de l'installation de JAVA.)

Avec *Oui*, on obtient la possibilité d'indiquer un dossier vide ou un nouveau où une copie de la livraison SIARD sera enregistrée. Après l'installation réussie, *SiardEdit* peut être lancé depuis le dossier choisi.

La fenêtre principale de *SiardEdit* apparaît ensuite, indépendamment du fait que *SiardEdit* ait été exécuté à partir du CD-ROM ou qu'il ait été installé sur le PC de l'utilisateur.

7.2 Fenêtre principale

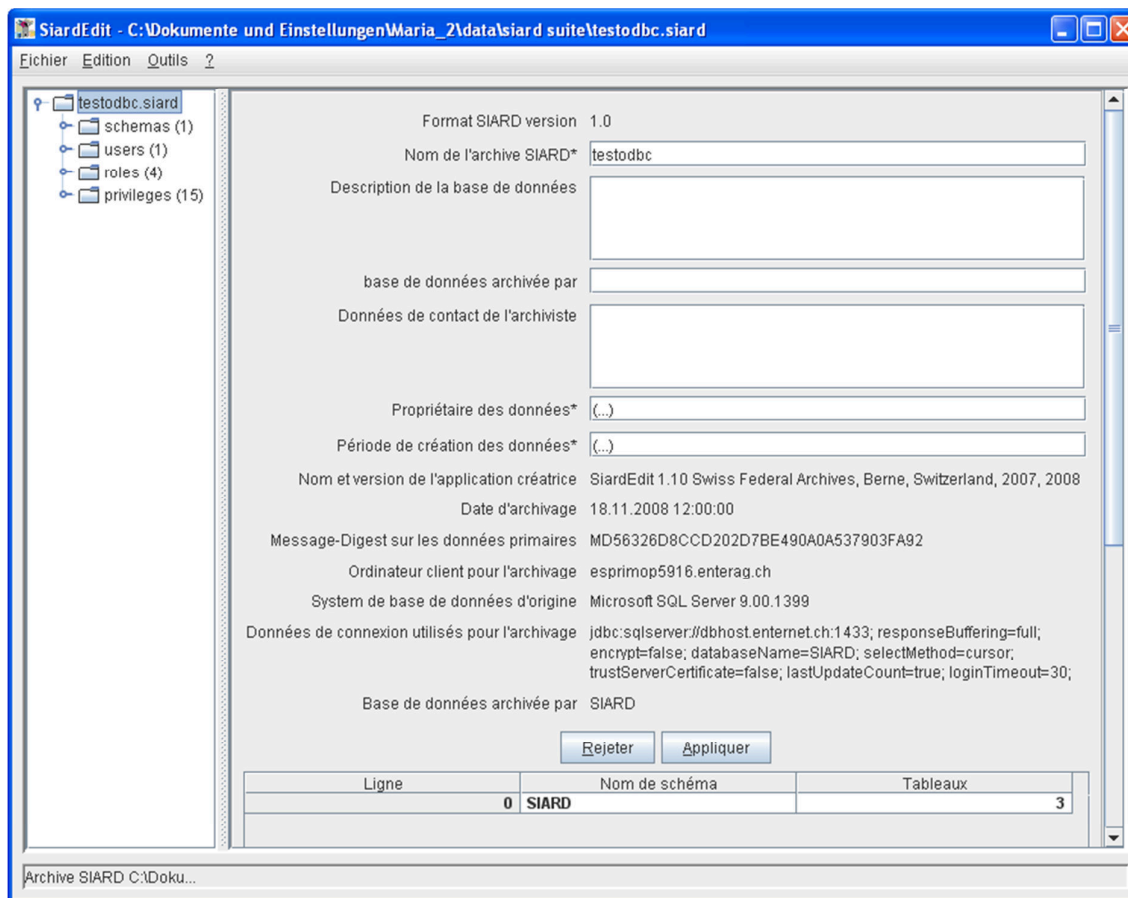
La fenêtre principale se compose de menus (en haut), d'une arborescence de navigation (à gauche), d'un contenu (à droite) et d'une ligne d'état (en bas).



On peut déplacer de manière flexible la limite entre la navigation et le contenu. La grandeur de la fenêtre entière peut être également modifiée jusqu'à une grandeur minimale donnée.

SIARD Suite

Si un fichier SIARD est chargé dans SiardEdit, la fenêtre principale est affichée comme suit:



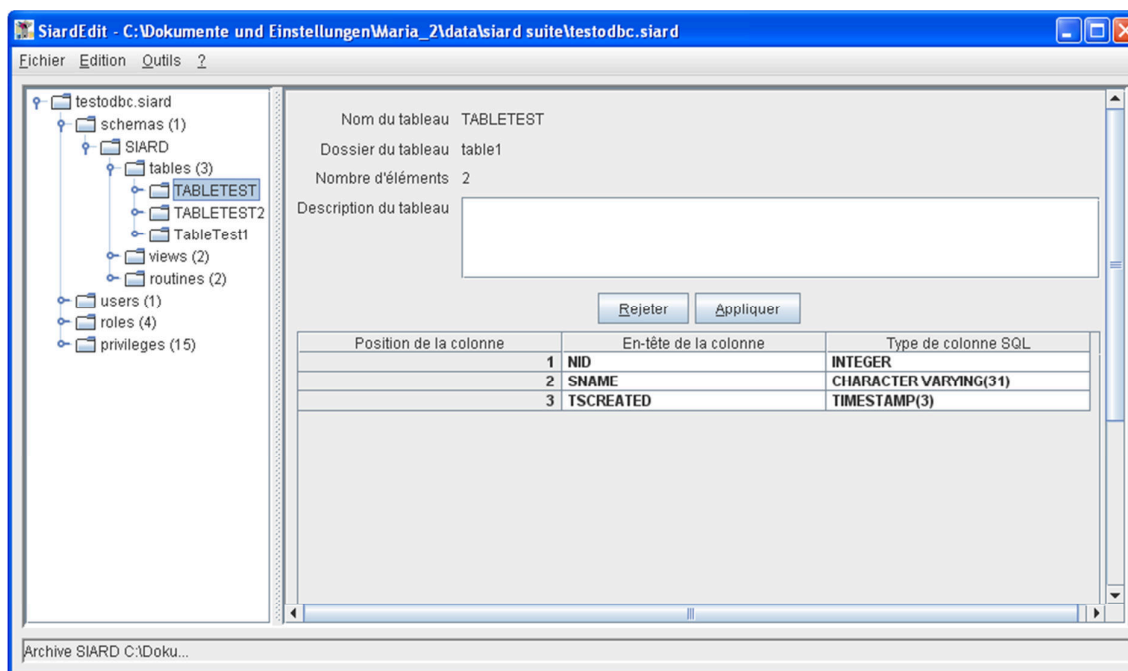
Dans la partie de gauche, on peut naviguer dans l'arborescence meta. Dans la partie de droite, on peut saisir en haut des métadonnées modifiables appartenant à l'objet de la base de données choisie à gauche, ou les modifier.

Le bouton *Appliquer* sert à la transmission des saisies dans les métadonnées du fichier SIARD momentanément ouvert. Si on confirme avec *Rejeter*, toutes les modifications seront annulées depuis la dernière reprise.

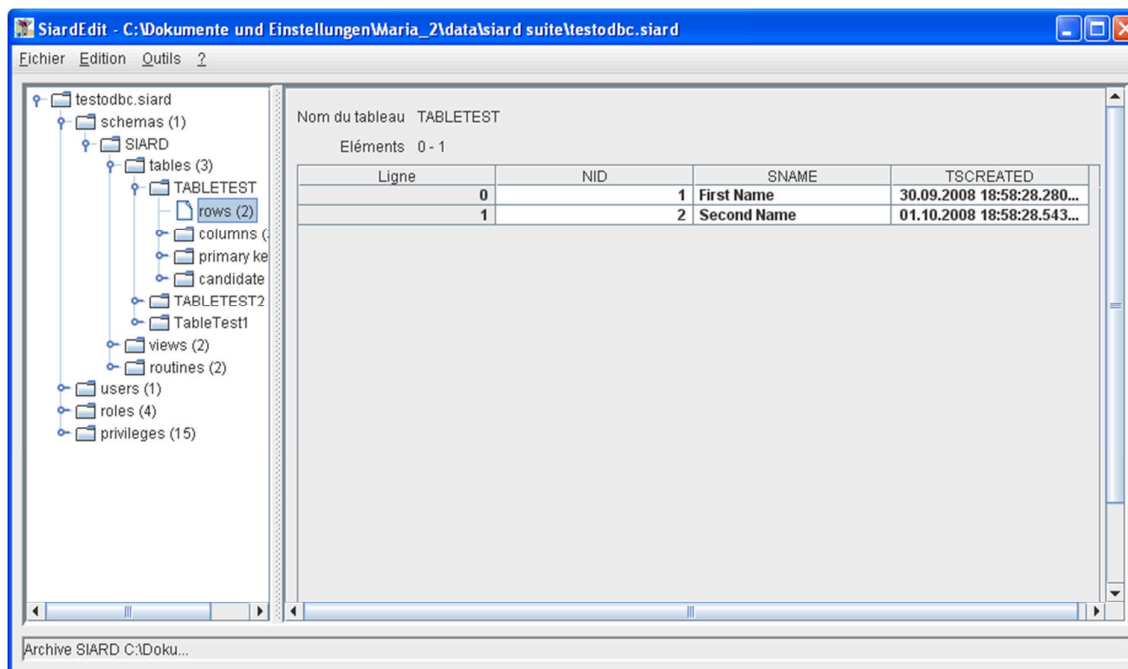
Sous ces métadonnées, un tableau des sous objets les plus importants est affiché. Les colonnes de ce tableau peuvent être triées en cliquant sur l'en-tête des colonnes. En effectuant un double clic sur une ligne, le sous objet correspondant est sélectionné.

Comme les tableaux dans un schéma ou les colonnes dans un tableau ne sont originellement pas triés dans un ordre spécifique, mais normalement triés par *SiardEdit* par ordre alphabétique, cette fonction de tri est très utile pour travailler avec de grosses bases de données.

SIARD Suite



Si, dans cette arborescence, on navigue sur le niveau tableau, on peut accéder aux données primaires en choisissant l'objet *rows* sous le tableau.



Pendant l'affichage des données primaires, *SiardEdit* modifie légèrement son comportement. En cliquant sur l'en-tête d'une colonne, tout le tableau sera trié d'après cette colonne. Cela peut nécessiter un certain temps pour de grands tableaux.

SIARD Suite

Pour permettre de travailler avec de très grands tableaux, *SiardEdit* indique seulement n-ten(n une puissance de 50) enregistrement afin d'empêcher l'affichage d'un nombre impossible de lignes. En effectuant un double clic sur une ligne, les enregistrements contenus entre cette ligne et la ligne suivante sont affichés jusqu'à ce que le niveau le plus bas soit atteint, où tous les enregistrements sont affichés les uns derrière les autres.

En effectuant un double clic sur un texte, un chiffre ou un champ de date, le contenu des données est affiché dans un éditeur de texte. En effectuant un double clic sur un champ binaire, le contenu est affiché dans un éditeur binaire. Dans *Extras / Option*, il est possible de définir quelles applications doivent être utilisées pour l'affichage des champs. On peut aussi déplacer les limites des colonnes afin de rendre complètement visibles les données en chiffres ou de dates.

Si on sélectionne une ligne du tableau en cliquant dessus, la couleur en est modifiée. En sélectionnant le menu *Edition / Copier (Ctrl-C)*, on peut enregistrer la ligne dans le presse-papier. Si on sélectionne le menu *Edition / Tout copier (Ctrl-A)*, on peut copier tout le tableau affiché dans le presse-papier. Le contenu du presse-papier peut être ensuite collé par ex. dans un programme de calcul de tableaux pour un traitement consécutif.

7.3 Menus

Les menus sont expliqués dans les chapitres suivants.

7.3.1 *Fichier*

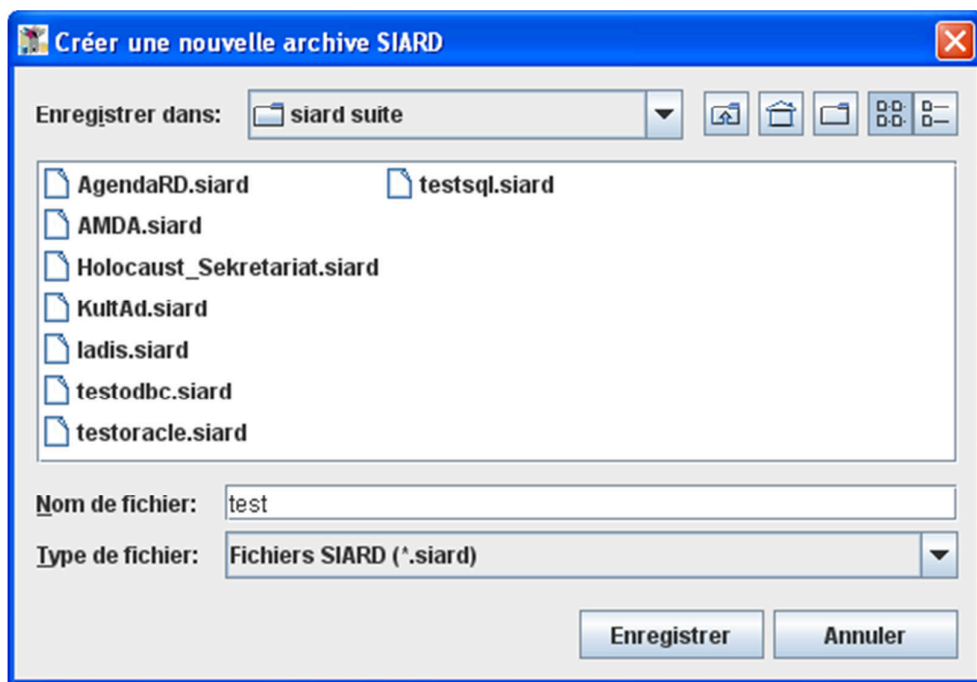
Nouvelle archive de base de données...

Cette fonction permet de télécharger une base de données et de l'enregistrer en fichier SIARD. Le fichier SIARD enregistré est ensuite affiché dans *SiardEdit*.

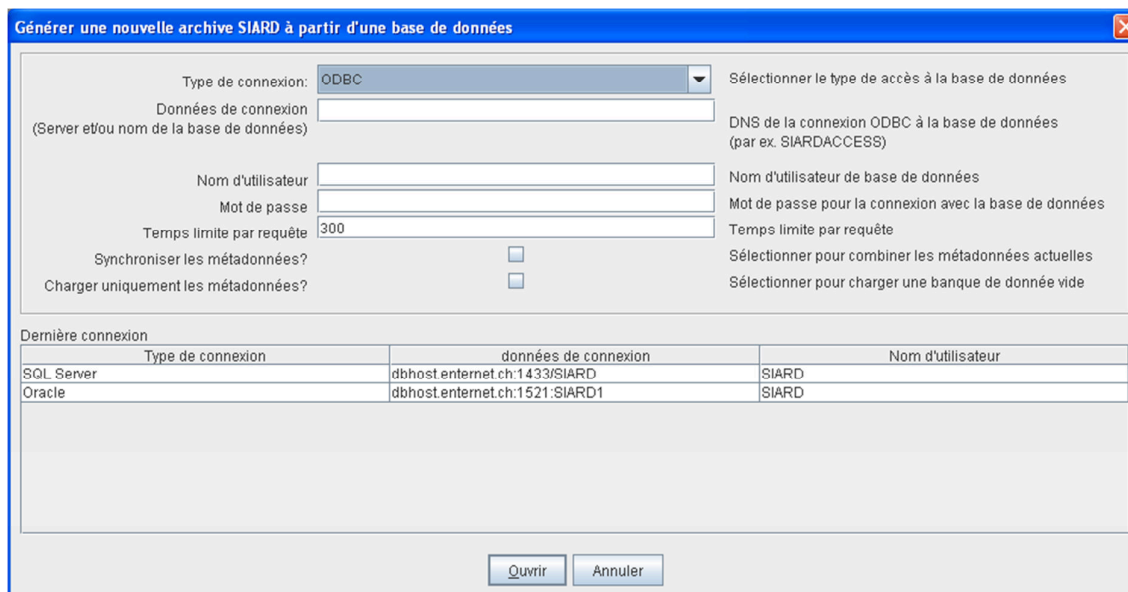
Cette action, en termes de fonction, est identique à l'application ligne de commande *SiardFromDb*. Dans la description de ce programme, on trouve des indications importantes, sur la manière de préparer les différents systèmes de bases de données pour cette opération.

En sélectionnant ce menu, on doit indiquer en premier sous quel nom le nouveau fichier SIARD à créer doit être enregistré.

SIARD Suite



Dans le dialogue *Générer une nouvelle archive SIARD à partir d'une base de données*, on doit saisir les données de connexion à la base de données devant être téléchargée:



SIARD Suite

Type de connexion

Sous type de connexion choisir:

- *ODBC* pour bases de données MS Access,
- *Oracle* pour bases de données Oracle
- *MySQL* pour bases de données MySQL ou
- *SQL Server* pour bases de données SQL Server.
- *DB/2* pour bases de données DB/2

Données de connexion

Les données de connexion sont différemment structurées selon le type de connexion.

ODBC

Il faut avant tout avoir configuré une connexion ODBC pour la base de données désirée, afin de pouvoir utiliser le type de connexion *ODBC*. La marche à suivre se trouve dans la documentation d'aide du système d'exploitation.

Dans les données de connexion, il faut saisir dans ce cas le DSN (Datasource Name) de la connexion ODBC.

Oracle

Ici, il faut saisir une étiquette de signes qui se compose du nom DNS (DNS = Domain Name System) du serveur de base de données, du port listener et du SID (System ID) de la base de données.

Ces données sont séparées par un double point. Exemple pour des données de connexion Oracle: *athlon700b.enterag.ch:1521:SIARD*.

Ici, *athlon700b.enterag.ch* est le nom DNS du server, 1521 le Port Listener de la base de données (valeur standard pour bases de données Oracle, pour autant que votre configuration n'ait pas été modifiée par l'administrateur de bases de données) et *SIARD* la SID de la base de données.

SQL Server

Ici, il faut saisir une étiquette de signes qui se compose du nom DNS (DNS = Domain Name System) du serveur de base de données, du port listener et du nom de la base de données. Serveur et port listener sont séparés par un double point. Le nom de la base de données est séparé par un trait oblique.

Exemple pour données de connexion pour SQL Server: *athlon800b.enterag.ch:1433/siard*.

Ici, *athlon800b.enterag.ch* est le nom DNS du server, 1433 le Port Listener de la base de données (valeur standard pour bases de données SQL Server, pour autant que votre configuration n'ait pas été modifiée par l'administrateur de bases de données) et *siard* le nom de la base de données.

MySQL

Il faut ici saisir une chaîne de signes qui se compose du nom DNS (DNS = Domain Name System) du serveur de la base de données, du port listener et du nom de la base de données (nom du schéma). Le serveur et le port sont séparés par deux points. Le nom de la base de données resp. le nom du schéma est séparé par une barre oblique.

Exemple pour données de connexion pour MySQL : *asusp5b.enterag.ch:3306/SIARD*.

SIARD Suite

Ici, *asusp5b.enterag.ch* est le nom DNS du Server, 3306 le Port Listener de la base de données (valeur standard pour bases de données MySQL, pour autant que votre configuration n'ait pas été modifiée par l'administrateur de bases de données) et *siard* le nom de la base de données.

DB/2

Il faut ici saisir une chaîne de signes qui se compose du nom DNS (DNS = Domain Name System) du server de la base de données, du port listener et du nom de la base de données (nom du schéma). Le server et le port sont séparés par deux points. Le nom de la base de données resp. le nom du schéma est séparé par une barre oblique.

Exemple pour données de connexion pour DB/2: *vmu64.enterag.ch:50000/SIARD*.

Ici, *vmu64.enterag.ch* est le nom DNS du Server, 50000 le Port Listener de la base de données (valeur standard pour bases de données DB/2 du type Express-C, pour autant que votre configuration n'ait pas été modifiée par l'administrateur de bases de données) et *siard* le nom de la base de données.

Utilisateur base de données

Ici est indiqué l'utilisateur de la base de données pour lequel l'aperçu de la base de données doit être archivé. Attention: certains systèmes de bases de données ignorent minuscules et majuscules, d'autres pas!

Si seulement des parties d'une base de données doivent être archivées, il faut configurer un utilisateur d'archive et lui attribuer uniquement des droits de lecture sur les objets de bases de données qui devront être archivés.

Mot de passe base de données

Aussi bien l'utilisateur de bases de données que le mot de passe doivent être créés par l'administrateur de bases de données. Ils peuvent être demandés par celui-ci, s'ils ne sont pas connus. Attention: les anciens systèmes de bases de données ignorent minuscules et majuscules, mais pas les nouveaux!

Pour les bases de données MS Access, l'utilisateur s'appelle normalement *Admin* et le mot de passe n'existe pas (chaîne de caractères vide).

Charger uniquement des métadonnées?

En choisissant cette option, un fichier SIARD sans données primaires est créé. Ceci est utile si on veut d'abord obtenir un aperçu de la taille de la base de données. Dans un tel fichier SIARD vide, on peut déjà insérer des déclarations de métadonnées et les ajouter à celles-ci plus tard à l'aide de "Synchroniser", lors du téléchargement de la base de données complète.

Synchroniser?

Cette option apparaît seulement lorsqu'un autre fichier SIARD a été chargé dans la fenêtre principale. Si cette option est choisie, les métadonnées du fichier SIARD chargé auparavant seront copiées aussi bien que possible dans les métadonnées du fichier SIARD nouvellement chargé. (Cela signifie que les enregistrements dans des tableaux, des colonnes, etc. identiques dans les deux schémas de bases de données seront repris.)

Dernières connexions

Dans ce tableau, les dernières connexions établies avec succès par l'utilisateur sont affichées. Elles sont enregistrées dans le fichier *siard.properties* du dossier *java* dans le dossier home de l'utilisateur.

SIARD Suite

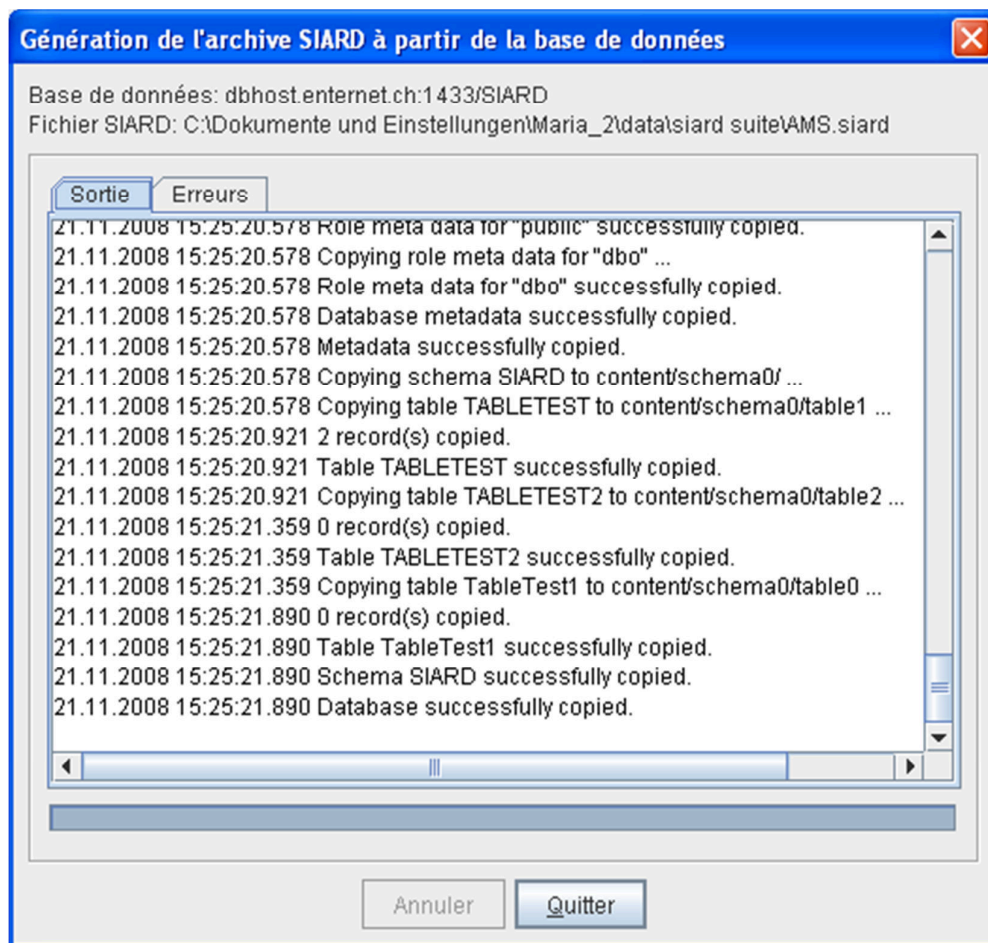
En cliquant sur une ligne de ce tableau, les données de connexion correspondantes seront remplies dans le dialogue. Seul le mot de passe reste vide et le curseur saute sur le champ du mot de passe.

Ouvrir

Ce bouton permet de charger la base de données dans le fichier SIARD choisi. Un protocole, qui est identique à celui donné par *SiardFromDb*, est affiché.

Annuler

Ce bouton permet d'annuler l'action.



A la fin du chargement, on peut encore marquer le protocole et le copier et ensuite, à l'aide d'un programme de traitement de texte, l'imprimer. Les indications d'heures à côté des différentes étapes servent à estimer le temps de chargement global de grosses bases de données.

En choisissant le bouton *Exit*, le fichier nouvellement chargé est ouvert dans la fenêtre principale.

Ouvrir...

En choisissant ce menu, on peut charger un fichier SIARD (ayant déjà été chargé).

SIARD Suite

Ses métadonnées apparaissent alors dans la fenêtre principale et peuvent être enregistrées après avoir été complétées et modifiées.

Enregistrer ... (Ctrl-S)

Ce menu permet d'enregistrer les métadonnées modifiées dans les fichiers SIARD ouverts.

Charger dans base de données ...

Ce menu permet de sélectionner la base des données, dans laquelle le contenu d'un fichier SIARD ouvert doit être chargé.

En termes de fonction, cette action est identique à l'application ligne de commande *SiardToDb*. Dans la description de *SiardToDb*, on trouve des indications importantes, comment préparer les différents systèmes de bases de données pour cette opération. Un utilisateur de bases de données adéquat doit être préalablement créé avec les droits nécessaires. Le chargement dans MS Access nécessite tout d'abord la création d'une base de données vide et un accès par ODBC.

En choisissant ce menu, le dialogue *Charger fichier SIARD dans base de données* va demander la saisie des données de connexion à la base de données dans laquelle les données doivent être chargées.

Charger le archive SIARD dans une base de données

Type de connexion: ODBC Sélectionner le type de accès à la base de données

Données de connexion (Server et/ou nom de la base de données) DNS de la connexion ODBC à la base de données (par ex. SIARDACCESS)

Nom d'utilisateur Norm d'utilisateur de base de données

Mot de passe Mot de passe pour la connexion avec la base de données

Temps limite par requête 300 Temps limite par requête

Dernière connexion

Type de connexion	données de connexion	Nom d'utilisateur
SQL Server	dbhost.enternet.ch:1433:SIARD	SIARD
Oracle	dbhost.enternet.ch:1521:SIARD1	SIARD

Ouvrir Annuler

Après l'établissement de la connexion avec la base de données upload, *SiardEdit* affiche un dialogue dans lequel on peut assigner le schéma de l'archive SIARD aux schémas dans la base de données upload.

Le chargement de contenus de bases de données est soumis aux restrictions des systèmes de bases de données dans lesquels les données sont chargées. Ainsi, MS Access ne supporte aucun schéma de bases de données. C'est pourquoi dans ces systèmes, tous les tableaux sont chargés dans la base de données sans schéma de séparation.

L'avantage du chargement de contenus de bases de données réside dans le fait que dans un système de bases de données, toute la puissance d'expression de SQL est disponible pour formuler des requêtes.

SIARD Suite

Exporter les métadonnées ...

Cette fonction permet d'exporter les métadonnées SIARD sous forme de fichiers XML.

En théorie, on peut aussi ouvrir le fichier SIARD avec *PKZIP* et exporter les métadonnées depuis le dossier *header*. Cependant, tous les utilisateurs ne disposent pas d'une licence *PKZIP*. Une telle licence serait toutefois nécessaire, car d'autres programmes ZIP (Info-ZIP, Microsoft Windows XP dossier ZIP, etc.) ne supportent pas (encore) le standard 64 bit. Comme les fichiers SIARD peuvent devenir très gros, il a été nécessaire d'utiliser ce standard 64 bit qui permet l'enregistrement de fichiers ZIP plus gros que 4 GB et contenant plus de 66'000 fichiers. A la place de *PKZIP*, on peut aussi utiliser l'application open source *Zip64File*⁸ pour décompresser les fichiers ZIP64.

Le fichier exporté *metadata.xml* peut être transformé en rapport sur les métadonnées de la base de données du fichier SIARD à l'aide de simples fichiers XSLT. En copiant le fichier *metadat.xsl* du dossier *doc* de l'installation SIARD dans le même répertoire que le fichier XML exporté *metadata.xml* et en l'ouvrant par exemple avec un des browsers les plus modernes, les métadonnées seront présentées dans un tel format de rapport.

La description de la structure des métadonnées et du format de fichier SIARD se trouve dans le document *SIARD Format 2.0_fr.pdf*, leur définition formelle dans les définitions de schémas XML *metadata.xsd* dans le dossier *doc* de l'installation SIARD.

Ouvrir les fichiers récents

On trouve ici les fichiers utilisés en dernier permettant un accès rapide.

Exit

Dans le menu *Fichier*, on trouve la fonction *Exit*. Cette dernière permet de terminer le programme.

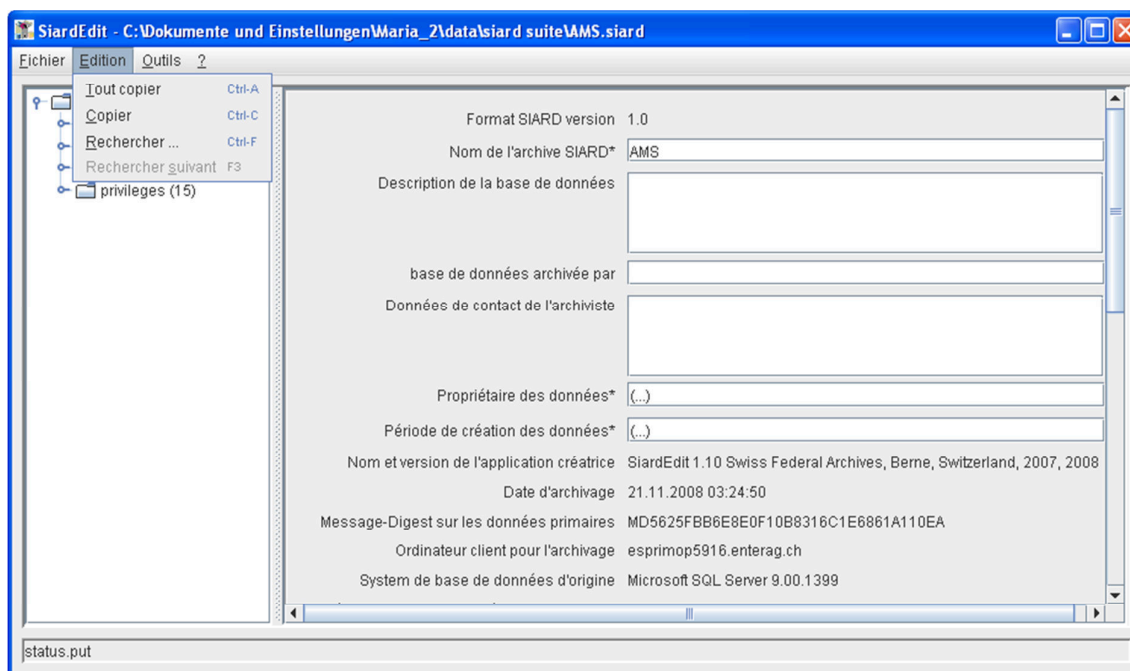
7.3.2 Edition

Copier tout (Ctrl-A)

Dans le menu *Edition* se trouve la fonction *Copier tout (Ctrl-A)*.

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

SIARD Suite



Avec cette fonction, on peut enregistrer le tableau en bas à droite en format texte séparé par des tabulateurs dans le presse-papier. Ensuite, on peut le coller de préférence dans un logiciel de calcul de tableaux pour un traitement ultérieur.

Copier (Ctrl-C)

Cette fonction permet de reprendre la ligne actuelle colorée en rose du tableau en bas à droite dans le presse-papier (avec tabulateurs comme séparateurs), afin de pouvoir l'insérer dans une autre application.

Rechercher ... (Ctrl-F)

Cette fonction permet de rechercher une chaîne de caractères dans les métadonnées.

Rechercher suivant (F3)

Cette fonction permet de poursuivre la recherche d'après une chaîne de caractères dans les métadonnées.

7.3.3 Extras

Installer

Dans le menu *Extras* on trouve la fonction *Installer* qui permet l'installation ultérieure de SIARD sur le PC.

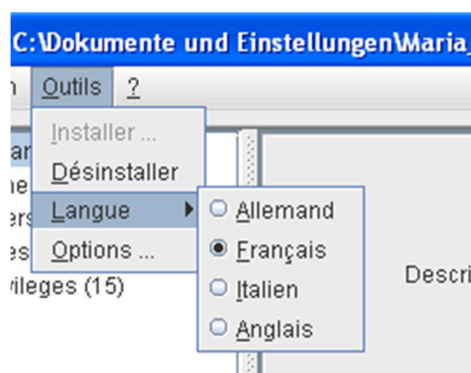
Désinstaller

Dans le menu *Extras*, on peut avec cette fonction supprimer l'installation locale de SIARD.

Langue

Dans le menu *Extras*, l'utilisateur peut au moyen de la fonction *Langue*, sélectionner la langue d'utilisateur désirée.

SIARD Suite



La fenêtre principale *SiardEdit* s'adapte de suite à la langue sélectionnée. En quittant le programme, la langue d'utilisateur est enregistrée dans le fichier `<home>/java/siard suite.properties`. `<home>` signifie sous LINUX le répertoire `$HOME` et sous Windows le dossier `C:\Documents and Settings\UserID\`. (`<UserID>` représente le nom sous lequel l'utilisateur est annoncé dans Windows.)

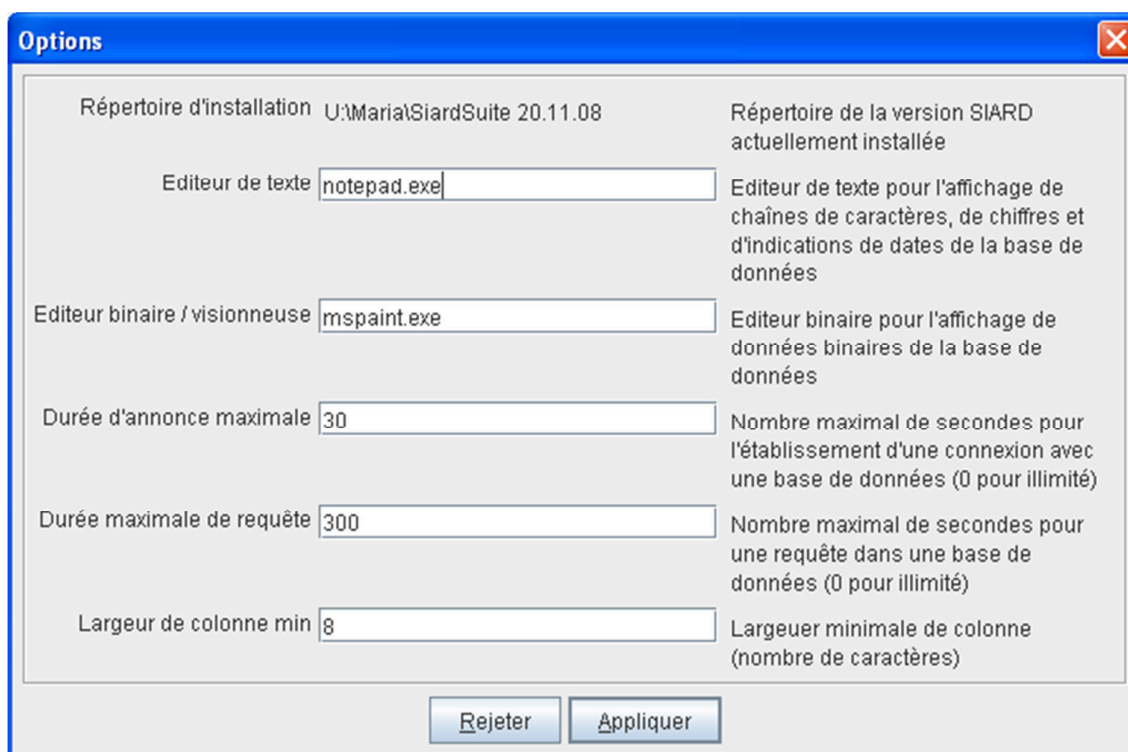
Les utilisateurs expérimentés peuvent aussi imposer un autre répertoire en indiquant lors du lancement de JAVA dans la ligne de commande `"-Duser.dir=<autreRépertoire>"`

Contrôle d'intégrité

Avec la fonction *Contrôle d'intégrité* du menu Extras, on peut vérifier que les données primaires n'ont pas été modifiées, la signature (MD5 ou SHA-1) a été calculée et enregistrée dans les métadonnées. Les données primaires ne sont "scellées" qu'une fois, immédiatement après le téléchargement réussi depuis une base de données avec une telle signature.

Options

Dans le menu *Extras*, on trouve la fonction *Options ...* qui permet plusieurs réglages spécifiques à l'utilisateur.



Les indications pour *Editeur texte* et *Editeur binaire* sont utilisées lorsqu'un double clic est effectué sur une cellule de données primaires du type chaîne de caractères ou du type données binaires. Le contenu de la cellule sera enregistré dans un fichier temporaire et ouvert au moyen de l'application indiquée. L'éditeur binaire doit particulièrement être modifié en fonction du type de données CLOB.

Le *Temps maximum d'annonce* détermine après combien de secondes une tentative d'annonce sera interrompue. Il ne doit être modifié que si l'on a des raisons de penser que le réseau ou la base de données réagit très lentement.

Le *Temps maximum de requête* ne devrait être réglé sur zéro que pour le chargement de très grosses bases de données parce que, dans ce cas, les requêtes peuvent durer plus longtemps que les 5 minutes indiquées dans les réglages de base.

La *Largeur minimale des colonnes* détermine l'affichage des tableaux dans la partie inférieure droite de la fenêtre principale.

7.3.4 ?

Aide

En cliquant sur ce menu, la table des matières de l'aide est affichée.

Info

Dans le menu ?, les dispositions relatives aux droits d'auteurs de *SiardEdit* sont affichées:



En outre, l'architecture du système d'exploitation et le runtime JAVA sont affichés ici.

7.3.5 *Aide contextuelle*

En principe, l'utilisateur dispose partout d'une aide contextuelle qui peut être appelée en pressant F1. L'aide affichée dans ce cas se rapporte à l'élément ou à la fenêtre dans lesquels des actions sont effectuées par le clavier ou la souris.

Si la fonction Aide est choisie depuis ?, une explication de cette fonction d'aide est affichée.

Dans le menu d'aide, il est possible d'appeler des pages d'explications depuis la table des matières ou de rechercher à l'aide de mots-clés. Une page affichée est en principe chaque fois synchronisée avec la table des matières.

8 SiardFromDb

Le programme *SiardFromDb* est un programme de ligne de commande qui extrait une base de données dans une archive SIARD. Avec *SiardFromDb*, il est possible de

- créer une archive SIARD "complète" (métadonnées et données primaires) basée sur la base de données (option "full"),
- créer une archive SIARD "vide" (sans données primaires) avec une définition schématique, basée sur une base de données (option "empty").

8.1 Commande

```
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>]
```

Comme <siardpath>, il faut saisir le répertoire où SIARD a été installé. Dans le sous-dossier bin du répertoire d'installation se trouve le fichier *SiardEdit.jar* avec la classe *ch.admin.bar.siard.SiardFromDb*, duquel *main()* ici sera exécuté avec la commande *java* (sous Windows, choisir de préférence *javaw*).

Lorsque l'option *-h (help)* est présente dans la ligne de commande, la syntaxe de commande est affichée.

Comme la commande contient un grand nombre de paramètres, on peut la simplifier avec un fichier de configuration dans lequel tous les paramètres seront enregistrés, à l'exception du mot de passe. Si ce fichier existe, la commande se réduit à:

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

Si tous les paramètres et en plus un fichier de configuration ont été indiqués, les paramètres du fichier de configuration seront complétés par ceux indiqués dans la ligne de commande. Ceci simplifie la répétition de téléchargements SIARD.

Lorsque l'option *-e (empty)* est indiquée, seules les métadonnées seront enregistrées.

8.2 Paramètres de commande

Paramètre	Signification
<data source>	ODBC, Oracle, "Server SQL", MySQL ou "DB/2"
<database name>	Nom de la base de données différemment structuré selon le type de connexion: ODBC: datasourcename (DSN) configuré de la connexion ODBC Oracle: par ex. dbhost.enternet.ch:1521:SIARD1 SQL Server: par ex. dbhost.enternet.ch:1433/SIARD MySQL: par ex. dbhost.enternet.ch:3306/SIARD

SIARD Suite

Paramètre	Signification
	DB/2: par ex. dbhost.enternet.ch:50000/SIARD
<database user>	UserId pour la session download
<database password>	Mot de passe pour la session download
<siard file>	Nom de fichier de l'archive SIARD à créer
<query timeout>	Timeout pour requêtes SQL en secondes (0 pour illimité)
<template file>	Fichier SIARD pouvant être utilisé comme modèle pour remplir les champs à saisir manuellement.

8.3 Remarques

Le programme *SiardFromDb* devrait être exécuté d'un "snapshot" de la base de données qui ne change pas pendant l'archivage.

L'archivage s'effectue complètement avec succès ou échoue complètement.

Pour de gros tableaux, il est recommandé de charger préalablement un fichier SIARD vide. Ainsi, on obtient un aperçu de toutes les métadonnées et des tailles des tableaux. Ces données aident à estimer le temps de download. En outre et pour de grands tableaux, il faudrait indiquer chaque fois l'option `-q 0` car il n'est pas possible d'estimer combien de secondes prendra une requête de taille pour un grand tableau.

La conversion des TIMES ou TIMESTAMPS dans la base de données dépend de la zone horaire locale. Lorsqu'on enregistre par exemple l'heure 15.30 à Zurich, l'heure UTC 14.30 (en hiver!) sera enregistrée dans le fichier XML. Si les heures inchangées dans la base de données doivent être interprétées comme heures UTC, il faut démarrer *SiardFromDb* avec l'option

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

8.4 Préparation de la base de données

Selon le type de base de données, Il faut préparer une base de données pour l'archivage de manière différente. Il faut être conscient que le choix du User-ID de la base de données pour le téléchargement est déterminant quant au volume de la base de données à archiver. Le fichier SIARD créé contiendra les objets de la base de données qui seront visibles pour l'utilisateur. Pour cette raison, ce n'est normalement pas une bonne idée d'utiliser pour le téléchargement, un administrateur base de données (*SYS* ou *SYSTEM* pour Oracle, *dbo* pour Server SQL, *mysql* ou *information_schema* pour MySQL, *db2inst1* pour DB/2 Express-C). Ces utilisateurs ont normalement trop de droit de lecture sur les objets système n'étant pas à archiver. Cependant, pour les bases de données MS Access, l'archivage du contenu le plus complet possible est souhaité. C'est pourquoi, dans ce cas, l'utilisation de *Admin* comme UserId de la base de données est recommandé.

Le fait que le choix du UserId est déterminant quant au volume de la base de données archivées peut être utilisé pour contrôler exactement le volume des données qui sont enregistrées dans le fichier SIARD: on crée un nouvel utilisateur d'archivage dans la base de données et on lui attribue un droit de lecture exactement pour les objets qui doivent être archivés.

SIARD Suite

8.4.1 Préparation MS Access

On peut ouvrir la plupart des bases de données MS Access avec UserId *Admin* et mot de passe " ". Pour obtenir un accès à des bases de données mieux protégées, il faut consulter la documentation du développeur.

Fichiers accdb et mdb

MS Access 2007 supporte deux formats de fichiers: l'"ancien" format de fichier avec la terminaison *.mdb* et le "nouveau" avec la terminaison *.accdb*. Les fichiers *accdb* supportent deux types de données – spécialement des champs plurivalents et des champs "annexe" (également plurivalents). En outre, les fichiers *accdb* interdisent l'accès aux tableaux de système et ainsi empêchent l'accès aux métadonnées enregistrées dedans via l'interface ODBC (et ainsi également via JDBC). Pour des fichiers *mdb*, on pouvait configurer un accès aux tableaux de système *MSysObjects*, *MSysRelationships* et *MSysQueries*. Cela permettait à *SIARD Suite* d'accéder à d'autres métadonnées qui ne sont pas accessibles via la classe JDBC *DatabaseMetaData*.

SIARD Suite traite ce nouvel état de choses de la manière suivante: s'il faut accéder en lecture aux tableaux *MSys*, votre information complémentaire est enregistrée dans l'archive SIARD. Autrement, aucune clé étrangère et aucun view ne peuvent être enregistrés. Pour des buts d'archivage, cette perte de métadonnées est souvent supportable, car aucune donnée primaire n'est concernée.

Le nouveau type de donnée *Annexe* dans les fichiers *accdb* est plus problématique. Microsoft a constaté que leur contenu ne peut être sélectionné d'aucune manière par l'interface ODBC (et ainsi JDBC). La version actuelle du format SIARD ne pourrait archiver raisonnablement ces champs BLOB plurivalents. Pour cette raison, *SIARD Suite* – comme les outils de migration Microsoft – reprend ces champs lors du téléchargement seulement comme champ texte qui contiennent des listes des noms de fichiers des annexes. Dans ce cas, le contenu des annexes est cependant perdu! Malheureusement, Microsoft empêche même *SIARD Suite* – et tous les autres clients ODBC et JDBC de détecter cette perte. Il incombe ainsi à l'archiviste de prendre les mesures nécessaires à ce que les annexes soient archivées séparément dans des fichiers *accdb*.

Si des fichiers *mdb* doivent être archivés, l'archiviste peut choisir s'il veut archiver seulement les métadonnées qui ont été archivées dans un fichier *accdb* – sans clé étrangère ou view –, ou s'il désire préparer le fichier *mdb* de telle manière que les tableaux système soient accessibles pour l'utilisateur *Admin*.

Questions d'architecture (32-bit / 64-bit)

Lorsque *SIARD Suite* est lancé avec un runtime JAVA 32-bit, seul un accès au pilote ODBC 32-bit est possible. Si MS Access 64-bit est installé sous Windows 64-bit, il ne faut configurer que des accès ODBC 64-bit avec l'administrateur standard ODBC (*C:<Windows\system32\odbcad32.exe*). Ceux-ci ne sont accessibles pour *SIARD Suite* que si il a été lancé avec un runtime JAVA 64-bit. Si l'on veut configurer un accès ODBC 32-bit, il faut utiliser l'administrateur ODBC 32-bit (*C:<Windows\SysWOW64\odbcad32.exe*). Ces accès ODBC 32-bit ne sont possibles pour *SIARD Suite* que s'il est lancé avec un runtime JAVA 32-bit.

Les fichiers batch *siardedit.cmd*, *siardfrommdb.cmd* et *siardtodb.cmd* (ainsi que le raccourci placé par l'installation sur le bureau) utilisent le runtime JAVA 64-bit, pour autant qu'il ait été installé. Si, en même temps, un runtime JAVA 32-bit est disponible, on peut lancer *SIARD Suite* manuellement avec celui-ci. Pour ce faire, il faut appeler les fichiers batch à l'aide du paramètre facultatif */32*. Si des problèmes surviennent lors du téléchargement d'un

SIARD Suite

fichier MS Access avec ODBC 64-bit, il vaut la peine d'essayer la même chose avec ODBC 32-bit. Actuellement, le pilote ODBC 64-bit comporte encore beaucoup d'erreurs.

Rendre accessible des tableaux système de fichiers mdb

1. Rendre lisibles les tableaux système *MSysObjects*, *MSysRelationships* et *MSysQueries* pour l'utilisateur de l'archive (la plupart du temps *Admin*).

MS Access avant 2007:

Dans ce but, rendre visibles les objets système sous *Outils / Options / Affichage* et après rendre lisibles pour l'utilisateur *Admin* les tableaux *MSysObjects*, *MSysRelationships* et *MSysQueries* sous *Outils / Sécurité / Gestion des utilisateurs et des groupes*.)

MS Access 2007:

Rendre visibles les objets système sous *Boutons Office / Options Access / Base de donnée active / Navigation.../ afficher les objets système* et ensuite rendre lisibles pour l'utilisateur *Admin* les tableaux *MSysObjects*, *MSysRelationships* und *MSysQueries* sous *Outils base de données / Administrer / Utilisateurs et autorisations / Autorisations Utilisateurs et groupes*. (Tout le groupe *Outils base de données / Administrer / Utilisateurs et autorisations* manque si le fichier ouvert est un fichier accdb.)

2. Configurer une connexion ODBC qui renvoie au fichier MDB (voir mot-clé ODBC dans l'aide Windows sous *Démarrer / Aide* pour les instructions sur la manière de configurer une connexion ODBC sous *Démarrer / Panneau de configuration / Outils d'administration / Sources de données (ODBC)*) et indiquer *ExtendedAnsiSQL=1* dans les propriétés étendues (sous *Avancé...*). La connexion ODBC peut être configurée comme *Utilisateur DNS* par de non administrateurs.

Configuration automatique et conversion

Un script (source ouverte) de commande *accesstosiard.js*⁹, ayant été développé comme complément de SIARD, assume tous ces travaux de configuration à la place de l'utilisateur et convertit simplement un fichier ACCDB ou MDB en fichier SIARD.

8.4.2 Préparation Oracle

Comme administrateur bases de données (DBA), on configurera une base de données Oracle de préférence pour l'exportation en créant un nouvel utilisateur d'archivage et en lui attribuant les droits de lecture sur tous les objets de bases de données à exporter. A l'aide de cette nouvelle identification utilisateur, on peut vérifier au moyen d'exemples de requêtes, si toutes les références nécessaires sont visibles. Il est recommandé de commenter et de documenter les exemples de requêtes pour les futurs utilisateurs de l'archivage.

Dépannage Oracle

Afin de réussir avec SIARD une connexion à la base de données Oracle désirée, les autorisations, pare feux, etc. doivent être configurés de telle manière que les commandes Oracle suivantes puissent établir une connexion:

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

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

SIARD Suite

Ensuite, dans *SIARD Suite*, le nom de la base de données `<hostname>:<port>:<service>` est à utiliser.

8.4.3 Préparation Serveur SQL

Comme administrateur bases de données (DBA), on configurera une base de données Server SQL de préférence pour l'exportation en créant un nouvel utilisateur d'archivage et en lui attribuant les droits de lecture sur tous les objets de bases de données à exporter.

Si on n'attribue pas de propriétés de détenteur des schémas à archiver à l'utilisateur de l'archive, celui-ci ne peut pas lire les privilèges existants et l'information enregistrée sur l'utilisateur et les rôles peut être incomplète.

Si l'utilisateur de l'archive obtient des droits SELECT, le schéma "dbo" – le plus souvent non souhaité – sera aussi automatiquement archivé. Les meilleures réglages pour l'utilisateur de l'archive sont donc:

Sous permissions base de données, n'attribuer que le droit CONNECT. Sous les propriétés utilisateur, définir l'utilisateur de l'archive comme détenteur de tous les schémas à archiver. (Attention, ce faisant, les anciens détenteurs seront "dépossédés". Il faut éventuellement annuler ce réglage après l'archivage.)

Dépannage Serveur SQL

Afin de réussir avec SIARD une connexion à la base de données SQL Server désirée, les autorisations, pare feux, etc. doivent être configurés de telle manière que la commande SQL Server suivante puisse établir une connexion:

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

Ensuite, dans *SIARD Suite*, le nom de la base de données `<hostname>:<port>/<database>` est à utiliser.

8.4.4 Préparation MySQL

Comme administrateur bases de données (DBA), on configurera une base de données MySQL de préférence pour l'exportation en créant un nouvel utilisateur d'archivage et en lui attribuant les droits de lecture sur tous les objets de bases de données à exporter.

```
CREATE USER '<USER>' IDENTIFIED BY '<PASSWORD>';
```

```
GRANT SELECT ON <SCHEMA>.* TO '<USER>';
```

(MySQL traite base de données et schéma comme termes échangeables. Au lieu de <SCHEMA>, il faut saisir ici le nom de la base de données.)

A l'aide de cette nouvelle identification utilisateur, on vérifie, au moyen d'exemples de requêtes, si toutes les références importantes sont visibles. Il est recommandé de commenter et de documenter les exemples de requêtes pour les futurs utilisateurs de l'archivage.

Dépannage MySQL

Afin de réussir avec SIARD une connexion à la base de données MySQL désirée, les autorisations, pare feux, etc. doivent être configurés de telle manière que la commandes MySQL suivante puisse établir une connexion:

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

Ensuite, dans *SIARD Suite*, le nom de la base de données `<hostname>:<port>/<database>` est à utiliser. (Attention: pas d'espace entre `-p` et `<password>`!)

Pour comprendre les droits d'accès d'un utilisateur d'une base de données MySQL, il faut considérer quatre tableaux dans INFORMATION_SCHEMA: USER_PRIVILEGES (droits

SIARD Suite

d'accès globaux), SCHEMA_PRIVILEGES (droits d'accès aux schémas), TABLE_PRIVILEGES (droits d'accès aux tableaux) et COLUMN_PRIVILEGES (droits d'accès aux colonnes).

8.4.5 **Préparation DB/2**

DB/2 ne reconnaît que les utilisateurs OS comme utilisateurs. Il faut donc configurer un utilisateur d'archive approprié et doter celui-ci avec GRANT ON DATABASE de tous les droits aux bases de données à archiver.

Attention: sous LINUX, l'identification utilisateur ne doit contenir que des minuscules. Elle doit être de huit caractères au maximum!

Dépannage DB/2

Le client DB/2 doit être installé sur le poste de travail. Sous Windows, il faut alors exécuter *db2cmd* afin d'obtenir l'accès à l'environnement des lignes de commandes db2. Sous LINUX, on obtient l'accès directement.

Par le biais de l'environnement des lignes de commandes db2, il faut configurer un nœud TCP/IP à l'aide de CATALOG et ensuite établir une connexion avec CONNECT.

Pour ce faire, il faut démarrer la fenêtre de lignes de commandes DB/2 ou le processeur de lignes de commandes db2 et effectuer les saisies comme suit:

a) LIST NODE DIRECTORY

Ceci devrait lister les nœuds enregistré sur le client. S'il n'existe pas de nœud pour l'instance DB/2 désirée, Il faut d'abord en créer une avec la commande:

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

Ici, il faut prendre de préférence comme <nodename> la première partie du nom DNS du server, aussi décrit comme Hostname. Le nom complet DNS du Serveur est à indiquer comme <dnsname>. Comme <port>, le numéro de port sur lequel les requêtes Listener DB/2 sont attendues (standard: 50000). Si, à présent, on exécute encore une fois la commande LIST NODE DIRECTORY, le nouveau <nodename> doit apparaître.

b) ATTACH TO <nodename> USER <siarduser>

Ici, le <nodename> est le nom du nœud qui définit l'instance DB/2 sur ce client. Comme <siarduser>, il faut saisir l'identification de l'utilisateur de l'archive de la base de données pour cette instance. Avec la commande ATTACH, le mot de passe de cet utilisateur est demandé.

c) LIST DATABASE DIRECTORY

Avec ceci, toutes les bases de données de l'instance DB/2, à laquelle on s'est connecté, sont affichées

d) CONNECT TO <database> USER <siarduser>

Comme <siarduser>, il faut saisir l'identification de l'utilisateur de l'archive de la base de données. Avec la commande CONNECT, le mot de passe de cet utilisateur est demandé. Ainsi, les commandes (SQL Statements) suivantes se rapportent à cette base de données.

Si ceci a été effectué avec succès, on peut lancer le téléchargement avec *SIARD Suite* au moyen de <dnsname>:<port>/<database> et l'utilisateur de l'archive avec <siarduser>.

9 SiardToDb

Le programme *SiardToDb* est un programme de ligne de commande qui charge une archive SIARD dans une instance de base de données pour des recherches.

9.1 Commande

```
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>]*
```

Comme <siardpath>, il faut saisir le répertoire où SIARD a été installé. Dans le sous-dossier bin du répertoire d'installation se trouve le fichier *SiardEdit.jar* avec la classe *ch.admin.bar.siard.SiardFromDb*, duquel *main()* ici sera exécuté avec la commande *java* (sous Windows, choisir de préférence *javaw*).

Lorsque l'option *-h* (*help*) est présente dans la ligne de commande, la syntaxe de commande est affichée.

Comme la commande contient un grand nombre de paramètres, on peut la simplifier avec un fichier de configuration dans lequel tous les paramètres seront enregistrés, à l'exception du mot de passe. Si ce fichier existe, la commande se réduit à:

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

Si tous les paramètres et en plus un fichier de configuration ont été indiqués, les paramètres du fichier de configuration seront complétés par ceux indiqués dans la ligne de commande. Ceci simplifie la répétition de téléchargements SIARD.

9.2 Paramètres de commande

Paramètre	Signification
<data source>	ODBC, Oracle, "SQL_Server", MySQL ou "DB/2"
<database name>	Nom de la base de données différemment structuré selon le type de connexion: ODBC: datasource name (DNS) configuré de la connexion ODBC Oracle: par ex. dbhost.enternet.ch:1521:SIARD1 SQL Server: par ex. dbhost.enternet.ch:1433/SIARD MySQL: par ex. dbhost.enternet.ch:3306/SIARD DB/2: par ex. dbhost.enternet.ch:50000/SIARD
<database user>	UserId pour la session upload
<database password>	Mot de passe pour la session upload
<siard file>	Nom de fichier de l'archive SIARD à charger
<query timeout>	Timeout pour requêtes SQL en secondes (0 pour illimité)
<schema>	Nom de schéma dans SIARD Archiv (à indiquer aussi souvent que désiré suivi de <mappedschema>)

<i>Paramètre</i>	<i>Signification</i>
<mappedschema>	Nom de schéma à utiliser pour le téléchargement. Son nom propre est utilisé pour tous les schémas pour lesquels aucun autre nom de schéma n'a été attribué.

9.3 Remarques

Comme les bases de données plus anciennes ne sont pas conformes à SQL:1999, une configuration manuelle fastidieuse ne peut être évitée pour la préparation à l'upload. Sous MS Access, il n'existe pas d'objets schémas. Sous SQL Server et Oracle 8 et 9, les utilisateurs et les objets schémas ne sont pas séparés.

Lors du chargement, seuls les tableaux sont créés. Tous les autres objets de la base de données ne sont pas générés. En outre, d'éventuelles déductions sont faites lors du chargement. Avec MS Access, tous les tableaux se retrouvent dans la même MDB. Avec Oracle, tous les noms plus longs que 30 caractères sont raccourcis. Pour éviter des collisions, les noms des tableaux et des colonnes sont complétés par un compteur. (Ainsi pour un tableau nommé "un nom de tableau trop long pour Oracle", on obtiendra "Un nom de tableau trop long poi".)

Où la précision maximale et le nombre maximal de décimales (par ex. dans MS Access) sont plus petits que nécessaires, les valeurs seront chargées avec une plus petite précision. SIARD s'appuie aussi bien que possible sur le système de bases de données visé. Si on consulte les métadonnées de la base de données avec *SiardEdit*, il sera possible d'attribuer les descriptions et les valeurs correctes.

La conversion de TIMES et TIMESTAMPS dans la base de données dépend de la zone horaire locale. Si par exemple l'heure (UTC) 14.30 est enregistrée dans le fichier SIARD Zurich, l'heure locale 15.30 (en hiver!) sera chargée dans la base de données. Si les heures doivent être téléchargées non modifiées comme heures UTC dans la base de données, Il faut démarrer *SiardToDb* avec l'option:

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

9.4 Préparation de la base de données

Selon le type de bases de données, il faut préparer la base de données pour l'upload de manière différente.

9.4.1 Préparation MS Access

Pour charger un fichier SIARD dans une base de données MS Access, il faut procéder comme suit:

1. Créer avec MS Access une mdb vide ou un fichier accdb.

Configurer une connexion ODBC pour le fichier de téléchargement vide créé. (voir mot-clé ODBC dans l'aide Windows sous *Démarrer / Aide* pour des instructions sur la manière de configurer une connexion ODBC sous *Démarrer / Panneau de configuration / Outils d'administration / Sources de données (ODBC)*) et indiquer *ExtendedAnsiSQL=1* dans les propriétés étendues (sous *Avancé...*). La connexion ODBC peut être configurée comme *Utilisateur DNS* par de non administrateurs.

SIARD Suite

Questions d'architecture (32-bit / 64-bit)

Lorsque *SIARD Suite* est lancé avec un runtime JAVA 32-bit, seul un accès au pilote ODBC 32-bit est possible. Si MS Access 64-bit est installé sous Windows 64-bit, il ne faut configurer que des accès ODBC 64-bit avec l'administrateur standard ODBC (C:<Windows\system32\odbcad32.exe). Ceux-ci ne sont accessibles pour *SIARD Suite* que s'il a été lancé avec un runtime JAVA 64-bit. Si l'on veut configurer un accès ODBC 32-bit, il faut utiliser l'administrateur ODBC 32-bit (C:<Windows\SysWOW64\odbcad32.exe). Ces accès ODBC 32-bit ne sont possibles pour *SIARD Suite* que s'il est lancé avec un runtime JAVA 32-bit.

Les fichiers batch *siardedit.cmd*, *siardfromdb.cmd* et *siardtodb.cmd* (ainsi que le raccourci placé par l'installation sur le bureau) utilisent le runtime JAVA 64-bit, pour autant qu'il ait été installé. Si, en même temps, un runtime JAVA 32-bit est disponible, on peut lancer *SIARD Suite* manuellement dans celui-ci. Pour ce faire, il faut appeler les fichiers batch à l'aide du paramètre facultatif /32. Si des problèmes surviennent lors du téléchargement d'un fichier MS Access avec ODBC 64-bit, il vaut la peine d'essayer la même chose avec ODBC 32-bit. Actuellement, le pilote ODBC 64-bit comporte encore beaucoup d'erreurs.

9.4.2 Préparation Oracle

Pour charger un fichier SIARD dans une base de données Oracle, il faut procéder comme suit:

1. Comme administrateur système, créer pour chaque schéma dans le fichier SIARD avec la commande CREATE USER un utilisateur (uniquement avec majuscules, car Oracle supporte les noms d'utilisateurs avec minuscules de manière erronée).
2. Attribuer les privilèges CONNECT et RESOURCE à un utilisateur d'archivage (par ex. un utilisateur de schéma).
3. Attribuer à l'utilisateur d'archivage tous les droits de création et d'écriture sur les schémas.

Le fichier SIARD ne contient plus d'un schéma que dans des cas d'exceptions. La plupart du temps, les instructions ci-dessus se réduisent à la ligne suivante:

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

Lors du chargement, de longues chaînes de caractères (par ex. VARCHAR(4000) sur des bases de données UTF-16 acceptant VARCHAR(2000) au maximum) sont transformées en CLOB. En outre, le paramètre de session NLS_LENGTH_SEMANTICS est placé sur CHAR (standard: BYTE) pour le chargement.

9.4.3 Préparation SQL Server

Pour charger un fichier SIARD dans une base de données SQL Server, il faut procéder comme suit:

1. Comme administrateur système, créer une base de données upload vide. (Dans Microsoft SQL Server Management Studio avec *New Database ...* sous *Databases* ou à l'aide de la commande CREATE DATABASE).
2. Pour l'utilisateur de l'archive, créer un *Login*. Il ne faut attribuer aucun privilège serveur à cet utilisateur de l'archive.
3. Seulement pour SQL Server 2000: Pour chaque schéma dans la base de données upload, créer un utilisateur pour lequel un *Login* est attribué.

SIARD Suite

4. La base de données upload attribue à l'utilisateur de l'archive au minimum les privilèges suivants:
 - a. Alter any role
 - b. Alter any schema
 - c. Alter any user
 - d. Alter
 - e. Connect
 - f. Create function
 - g. Create procedure
 - 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

Sous SQL Server 2000, les utilisateurs et les schémas étaient identiques. Sous SQL Server 2005/2010/2012 il faut exécuter pour chaque schéma la commande suivante:

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

Dans ce cas, il est suffisant d'attribuer à l'utilisateur de l'archive les privilèges Create pour la base de données.

9.4.4 *Préparation MySQL*

Pour charger un fichier SIARD dans une base de données MySQL, il faut procéder comme suit:

1. Comme administrateur système, créer à l'aide de la commande CREATE SCHEMA une base de données (= schéma) pour chaque schéma.
2. Attribuer à l'utilisateur de l'archive tous les droits de création et d'écriture sur tous les schémas.

Le fichier SIARD contient plus d'un schéma seulement en cas d'exceptions. La plupart du temps, les instructions ci-dessus se réduisent à ces lignes:

SIARD Suite

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

9.4.5 Préparation DB/2

Pour charger un fichier SIARD dans une base de données DB/2, il faut procéder comme suit:

Créer un utilisateur d'archive (par ex. *siarduser*) dans le système d'exploitation où le service DB/2 fonctionne.

Attention: s'il s'agit du système d'exploitation LINUX, l'id utilisateur de l'utilisateur d'archive ne doit comporter que des minuscules et ne pas dépasser 8 caractères!

Créer en tant qu'administrateur de base de données une base de données vide qui reprendra le contenu des fichiers SIARD.

Pour ce faire, il faut lancer la fenêtre de commandes DB/2 ou le processeur de lignes de commandes DB/2 et saisir ce qui suit:

a) LIST NODE DIRECTORY

Ceci devrait lister les nœuds enregistrés sur le client. Si aucun nœud n'existe pour l'instance DB/2 désirée, il faut d'abord en créer un avec la commande:

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

Ici, on prendra comme <nodename> de préférence la première partie du nom DNS du serveur qui est aussi défini comme hostname. Comme <dnsname>, il faut indiquer le nom complet DNS. Comme <port>, le numéro de port sur lequel le listener DB/2 attend les requêtes (standard: 50000). Si maintenant on exécute à nouveau la commande "LIST NODE DIRECTORY", le nouveau <nodename> doit apparaître.

b) ATTACH TO <nodename> USER <dbadm>

Ici <nodename> est le nom du nœud qui définit l'instance DB/2 sur ce client. Comme <dbadm>, il faut saisir l'id utilisateur d'un administrateur de base de données de cette instance. Avec la commande ATTACH, le mot de passe de cet administrateur de base de données sera demandé.

c) LIST DATABASE DIRECTORY

Avec cette commande, toutes les bases de données de l'instance DB/2 auxquelles on est connecté sont affichées.

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

On peut choisir librement le nom <newdatabase> pour la nouvelle base de données de l'instance DB/2. Il ne doit toutefois pas déjà exister. En outre, il ne doit pas comprendre plus de 8 caractères. Les majuscules ou minuscules sont ignorées.

L'utilisation du set UTF-8 est judicieuse car des chaînes de caractères avec trémas ne passent éventuellement pas dans les champs (mesurés en Byte pour DB/2). L'indication du territoire est obligatoire. On peut aussi prendre d'autres codes de pays ISO (par ex. CH, US,...).

e) CONNECT TO <newdatabase> USER <dbadm>

Comme <dbadm> il faut saisir l'id utilisateur d'un administrateur de base de données de cette instance. Avec la commande CONNECT, le mot de passe pour cet administrateur de

SIARD Suite

base de données est demandé. Ainsi, les autres commandes (statements SQL) se rapportent à la nouvelle base de données créée.

f) **CREATE SCHEMA <schema> AUTHORIZATION <siarduser>**

Pour chaque schéma dans le fichier SIARD. Ici, il faut indiquer comme <siarduser> l'id utilisateur créée comme utilisateur d'archive dans le système d'exploitation de l'instance DB/2.

g) Pour tester, il faut se connecter en tant que <siarduser> avec la nouvelle base de données.

CONNECT TO <newdatabase> USER <siarduser>

Ce faisant, le mot du <siarduser> est demandé.

Ces actions doivent évidemment être exécutées par un administrateur du serveur DB/2. Pour toutes les autres, seul l'utilisateur d'archive avec son mot de passe est nécessaire.

5.

10 Systèmes de bases de données supportés

Dans la *SIARD Suite*, les systèmes de bases de données „MS Access“, „Oracle“, „SQL Server“, "MySQL" et DB/2 sont supportés. Pour l'accès à MS Access, le pont JDBC/ODBC de SUN est utilisé, pour les autres systèmes de bases de données, des pilotes JDBC spécifiques aux bases de données sont utilisés.

En raison d'insuffisances des pilotes JDBC et ODBC, l'accès à d'autres systèmes de bases de données à l'aide d'ODBC n'est pas conseillé.

Comme les systèmes de bases de données supportent des types de données légèrement différents, la normalisation des types de données imposée par *SIARD Suite* est décrite dans un document séparé pour chaque système de base de données.

11 Protocole

Si des problèmes surviennent lors de l'utilisation de la *SIARD Suite*, il est utile d'enregistrer un protocole d'exécution. Un tel protocole devrait accompagner un message d'erreur à la maintenance. Pour enregistrer un protocole, il faut procéder comme suit:

Dans le répertoire *etc.* de l'installation SIARD se trouve le fichier *logging.properties*. Dans ce fichier, la journalisation au niveau INFO est déviée sur la console.. Pour dévier la journalisation dans un fichier, il faut enlever le signe de commentaire # devant.

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

En outre dans la ligne

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

il faut changer le nom de fichier après le signe d'égalité par le chemin d'accès complet du fichier de protocole désiré. Ce faisant, il faut tenir compte que les barres obliques inversées doivent être doublées. Cependant, sous Windows, on peut utiliser partout des barres obliques. Par exemple:

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

On peut aussi fixer le niveau de protocole (.level) dans ce fichier. Les valeurs admises sont: NONE, SEVERE, WARNING, INFO, CONFIG, FINE, FINER, FINEST, ALL. ALL crée la plupart du temps des enregistrements incompréhensibles pour des non développeurs. SEVERE crée seulement des informations sur des erreurs de système. INFO représente un réglage raisonnable.

12 Restrictions

Le format SIARD est soumis aux restrictions suivantes:

1. La grandeur du fichier SIARD ne peut dépasser 18'446'744'073'709'551'615 Bytes (ca. 18 ExaBytes) (Restriction ZIP64).
2. Le nombre de fichiers (tableaux et LOB) ne peut dépasser 4'294'967'295 (env. 4 milliards) (Restriction ZIP64).
3. L'ensemble des métadonnées de la base de données doit avoir place dans la mémoire JAVA.

Les deux premières restrictions ne devraient absolument pas représenter d'obstacles car les bases de données réelles ne deviendront encore longtemps pas si grandes.

Toutefois, la troisième restriction peut une fois survenir lors d'une constellation peu propice de la mémoire principale à disposition et de la complexité des métadonnées de la base de données. On peut échapper à cette restriction en lançant les applications SIARD sur une machine disposant suffisamment de mémoire principale (par ex. 4 GB) ou éventuellement en augmentant le Heap JAVA à l'aide du paramètre de ligne de commande JAVA „-Xmx2000m“ ou autre dans la commande JAVA.