

BIREME / PAHO / WHO

Centre d'information sur les sciences de la santé pour l'Amérique latine et les Caraïbes

**Principes élémentaires de bases de données CDS/ISIS:
initiation à l'utilisation de CISIS**

Version 3.x

Sao Paulo - 2007

Copyright © 2007 - BIREME / PAHO / WHO

Principes élémentaires de bases de données CDS/ISIS: initiation à l'utilisation de CISIS

Permission vous est donnée de copier, distribuer et/ou modifier ce document selon les termes de la Licence GNU Free Documentation License, Version 1.2 ou ultérieure publiée par la Free Software Foundation; sans les Sections Inaltérables, sans les les textes de première page de couverture ou les textes de dernière page de couverture. Une copie de cette Licence est incluse dans la section appelée "GNU Free Documentation License".

Notice de catalogage

BIREME / PAHO / WHO (Brésil)

Principes élémentaires de bases de données CDS/ISIS:
initiation à l'utilisation de CISIS. / BIREME / PAHO / WHO.
Sao Paulo : BIREME / PAHO / WHO, 2007.

30 p.

1.Manuel utilisateur. 2. Accès à l'information. 3.
Systèmes d'information. 4. Gestion de l'information. 5.
Santé publique. 6. Services de santé. I. BIREME II. Titre

Attention - Aucune mention faite dans ce document aux organismes, institutions, personnes ou produits, ne peut être considérée comme une approbation ou une recommandation donnée par BIREME / PAHO / WHO, et ne signifie pas qu'y ait préférence à d'autres de nature semblable, cités ou non.

BIREME / PAHO / WHO

Centre d'information sur les sciences de la santé pour l'Amérique latine et les Caraïbes

Rua Botucatu 862 V Clementino

Ce document a été produit avec le "Documents Conformation Methodology (NorDoc)" développé par BIREME.

Table des matières

Abréviations utilisées	IV
Préface	1
A propos de BIREME	1
La Bibliothèque virtuelle sur la santé (BVS)	2
Introduction	5
Bases de données CDS/ISIS	7
Maintenance de l'information dans la base de données	9
Jeux de caractères utilisés	10
Différences entre plate-formes	12
Initiation à l'utilisation de CISIS	13
MX l'utilitaire de CISIS	13
Autres commandes de CISIS	16
<i>crunchmf et crunchif</i>	16
<i>msrt</i>	16
<i>mxtb</i>	16
Aspects pratiques	19
Equivalence entre les commandes Linux et Windows	19
Commandes FTP les plus courantes	20
Conversion du jeu de caractères	20
Transfert de bases entre systèmes d'exploitation	21
Références bibliographiques	22
Glossaire	23

Abréviations utilisées

- ANSI. American National Standards Institute [Institut national américain de normalisation].
- ASCII. American Standard Code for Information Interchange [Code standard américain pour l'échange d'information].
- BIREME. Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud [Centre d'information sur les sciences de la santé pour l'Amérique latine et les Caraïbes].
- BVS. Biblioteca Virtual en Salud [Bibliothèque virtuelle sur la santé].
- CDS. Computerized Documentation System [Système de documentation informatisée].
- CP. Code Page [Page de codes].
- FST. Field Selection Table [Table de sélection des champs].
- FTP. File Transfer Protocol [Protocole de transfert de fichiers].

- **IFP.** Inverted File Pointer [Pointeur du fichier inversé].
- **ISIS.** Integrated Set of Information Systems [Ensemble intégré de systèmes d'information].
- **ISO.** International Organization for Standardization [Organisation internationale de normalisation].
- **LILACS.** Literatura Latinoamericana y del Caribe en Ciencias de la Salud [Littérature Latino-Américaine et des Caraïbes en sciences de la santé].
- **OMS.** Organización Mundial de la Salud [Organisation mondiale de la santé].
- **OPS.** Organización Panamericana de la Salud [Organisation panaméricaine de la santé].
- **UNESCO.** United Nations Educational, Scientific and Cultural Organization [Organisation des Nations Unies pour l'éducation, la science et la culture].

Préface

A propos de BIREME

Année après année, BIREME a rempli sa mission, à savoir être un centre spécialisé dans l'information scientifique et technique dans le domaine de la santé pour la région de l'Amérique latine et des Caraïbes. Fondée au Brésil en 1967, sous le nom de Bibliothèque régionale de médecine (d'où l'acronyme BIREME), elle a toujours répondu à la demande croissante en littérature scientifique à jour de la part des systèmes brésiliens de santé, des communautés de chercheurs, des professionnels et des étudiants du secteur de la santé. Puis en 1982, son nom changea en Centre d'information sur les sciences de la santé pour l'Amérique latine et les Caraïbes pour mieux exprimer ses fonctions orientées vers le renforcement et l'extension du flux d'information scientifique et technique sur la santé dans toute la région, mais elle a conservé son sigle.

Le travail en réseau, basé sur la décentralisation, le développement des capacités locales, le partage des ressources de l'information, le développement de produits et de services coopératifs, la conception des méthodologies communes, a toujours été le fondement du travail de coopération technique de BIREME. C'est de cette manière que le centre est devenu un modèle international qui promeut la qualification des professionnels de l'information au niveau managérial et

technique afin que ceux-ci adoptent les paradigmes de l'information et de la communication qui répondent le mieux aux besoins locaux.

Les principes les plus importants qui sont à l'origine du BIREME et qui sont les fondements de son existence sont :

- l'accès à l'information scientifique et technique sur la santé est essentiel pour le développement de la santé ;
- le besoin de développer la capacité des pays de l'Amérique latine et des Caraïbes à exploiter leurs sources d'information scientifique et technique sur la santé de manière coopérative et efficace ;
- la nécessité de promouvoir l'utilisation et de répondre aux demandes d'informations scientifique et technique sur la santé venant des gouvernements, des systèmes de santé, des établissements d'enseignement et de recherche.

BIREME, en tant que centre spécialisé de l'Organisation panaméricaine de la santé (OPS)/ Organisation mondiale de la santé (OMS), coordonne et réalise des activités de coopération technique en gestion de l'information et des connaissances scientifiques dans le but de renforcer et d'augmenter le flux d'information scientifique sur la santé au Brésil, dans les autres pays de l'Amérique latine et aux Caraïbes, comme condition principale pour le développement de la santé, y compris sa planification, gestion, promotion, recherche et éducation.

La convention qui soutient BIREME est renouvelée tous les cinq ans par les membres du Comité consultatif national de l'établissement (PAHO, ministère de santé brésilien, ministère brésilien de l'éducation et de la culture, secrétariat d'Etat à la santé de l'état de São Paulo, et université fédérale de São Paulo - Unifesp). Le dernier fournit l'infrastructure physique nécessaire pour l'établissement de l'institution.

En 2004 BIREME a pris la responsabilité de devenir un établissement basé sur la connaissance.

La Bibliothèque virtuelle sur la santé (BVS)

Avec l'apparition et le développement d'Internet comme moyen prédominant d'information et de communication, le modèle de coopération technique de BIREME a évolué depuis 1998 vers la construction et le développement de la

Bibliothèque virtuelle sur la santé (BVS) comme espace commun pour la convergence du travail coopératif des producteurs, intermédiaires et utilisateurs d'information. La BVS promeut le développement d'un réseau de sources d'information scientifique et technique avec l'accès universel à Internet. Pour la première fois il existe une réelle possibilité d'accès équitable à l'information sur la santé.

Pour BIREME, la Bibliothèque virtuelle sur la santé est un modèle pour la gestion de l'information et de la connaissance qui inclut la coopération et la convergence entre les établissements, les systèmes, les réseaux et les initiatives des producteurs, des intermédiaires et des utilisateurs dans l'exploitation des réseaux des sources d'information locales, nationales, régionales et internationales, en privilégiant l'accès libre et universel.

Aujourd'hui, tous les pays d'Amérique latine et des Caraïbes (Région) participent directement ou indirectement aux produits et services coopératifs offerts par la BVS, qui inclut plus de 1.000 établissements répartis dans plus de 30 pays.

La BVS est simulée dans un espace virtuel d'Internet formé par une collection ou un réseau de sources d'information sur la santé dans la Région. Les utilisateurs de niveaux et de lieux différents peuvent intervenir et naviguer dans l'espace d'une ou plusieurs sources d'information, indépendamment de l'endroit où ils se trouvent. Les sources d'information sont produites, mises à jour, stockées et exploitées sur Internet par des producteurs, des intégrateurs et des intermédiaires, de façon décentralisée, en suivant les méthodologies communes pour leur intégration dans la BVS.

La BVS organise l'information dans une structure qui intègre et relie des bases de données de référence, des annuaires spécialisés, des événements et des établissements, un catalogue des ressources de l'information disponibles sur Internet, des collections de textes intégraux avec entre autres la SciELO (Bibliothèque électronique scientifique en ligne) collection de périodiques scientifiques, des services sélectifs de diffusion de l'information, des sources d'information pour soutenir l'éducation et la prise de décision, des nouvelles, des listes de discussion et une aide aux communautés virtuelles. L'espace de la BVS est donc un réseau dynamique et décentralisé des sources d'information à partir

duquel il est possible de récupérer et extraire des 'informations pour soutenir les processus décisionnels dans le secteur de la santé.

La Bibliothèque virtuelle sur la santé peut être perçue comme une base partagée de la connaissance scientifique et technique dans le domaine de la santé qui est enregistrée, organisée et stockée dans le format électronique des pays de la Région, accessible de manière universelle sur Internet et compatibles avec les bases de données internationales.

Introduction

S'il y a une quantité significative d'informations, il est important de la préserver pour une consultation ultérieure.

Cette quantité d'informations ou de données peut, à juste titre, prétendre intégrer une banque de données.

Dans des termes plus traditionnels, une banque de données est considérée comme une collection de bases de données qui, en général, sont mises en rapport les unes avec les autres.

L'élément de base d'une base de données est l'enregistrement qui doit être considéré comme formant la plus petite partie d'information autonome. À son tour, cet enregistrement est composé d'éléments de données, qui sont stockés dans des champs et représentent une unité plus petite d'information, qui peut n'avoir aucune signification si elle est considérée de manière isolée. Le contenu d'un champ peut être composé d'éléments répartis dans des sous-champs. La figure 1 illustre cette description.

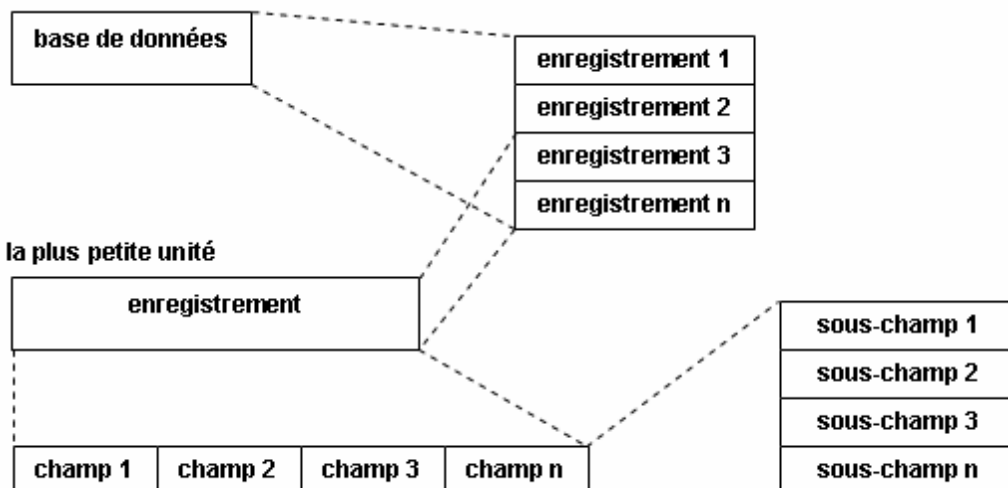


Figure 1 : Composition d'une banque de données

Supposons que chaque élément dans la figure 1 représente un élément catalogué, chaque champ correspond à un élément de données, comme l'auteur, le titre, les sujets. Quelques champs peuvent être composés de sous-champs ; par exemple le champ *auteur*, composé des sous-champs : nom, titre, institution à laquelle l'auteur appartient, département dans lequel il travaille, pour ne citer que quelques-unes des possibilités de sous-champs.

Il y a différents systèmes de gestion de banques et de bases de données ; actuellement, le système relationnel est reconnu comme "l'état de l'art" pour la gestion de banques de données. Il est basé sur le principe d'une suite de tables, entre lesquelles des relations logiques s'établissent.

Toutefois, en fonction de la nature de l'information, le modèle relationnel peut ne pas convenir lorsque, par exemple, les données peuvent ne pas faire l'objet de normalisation, ce qui serait contraire au modèle relationnel.

Pour le type d'information et de données que représente une référence bibliographique (principalement, de l'information textuelle), il faut adopter un modèle de banque de données plus souple, qui accepterait par exemple des enregistrements de longueur variable, des champs répétitifs dans un même enregistrement.

Bases de données CDS/ISIS

Une base de données CDS/ISIS a les caractéristiques suivantes :

- elle accepte des champs de longueur variable ;
- elle accepte la répétitivité des champs de données ;
- elle permet l'utilisation de sous-champs ;
- elle permet des relations entre des bases de données différentes ;
- elle utilise le fichier inversé comme moyen de récupération d'information.

Ainsi les éléments de données de chaque enregistrement occupent seulement l'espace (en bytes) dont ils ont besoin : il n'y a donc pas une longueur fixe pour tous les enregistrements et ceci empêche un gaspillage de l'espace de stockage.

Par conséquent puisqu'il est difficile de déterminer le début et la fin de chaque enregistrement dans la base de données, elle est techniquement composée de deux fichiers : l'un qui contient les données, avec en outre les informations de contrôle pour l'accès aux éléments de données, et l'autre qui contient les pointeurs qui signalent la position de chaque enregistrement. Le fichier avec les données et les contrôles d'accès est désigné avec l'extension .XRF (références croisées).

Celui qui est appelé "fichier inversé" met en place une structure de données appelée "arbre équilibré", dont la caractéristique est d'être extrêmement efficace

dans les opérations de recherche d'un élément aléatoire dans un contenu, et c'est ce qui fait la force de CDS/ISIS dans la récupération d'information. La solution technique pour mettre en oeuvre ce fichier inversé consiste en six fichiers ayant tous le même nom et avec les extensions : .CNT (contrôle); N01 (noeuds pour les termes courts); .L01 (feuilles pour les termes courts); .N02 (noeuds pour les termes longs); .L02 (feuilles pour les termes longs) et .IFP (pointeurs du fichier inversé).

Le concept de clés courtes et longues – non mentionné jusqu'à présent – a pour but d'optimiser l'efficacité de la base de données pour les opérations de récupération d'information. La taille des clés courtes est d'au plus 10 caractères, tandis que les clés longues ont de 11 à 30 caractères.

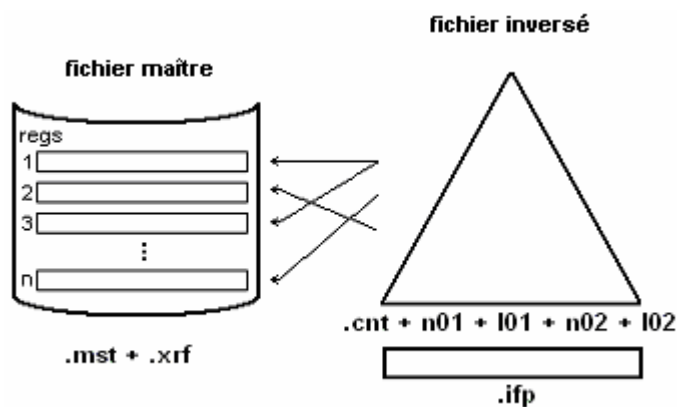


Figure 2 : Composition des bases de données CDS/ISIS

D'un point de vue fonctionnel, il y a deux éléments importants : le fichier maître qui contient toutes les informations de la base de données d'une part, et le fichier inversé qui contient la liste des termes qui peuvent être recherchés, avec leur localisation dans le fichier maître d'autre part.

On constitue ainsi une base de données d'informations à laquelle on peut accéder de façon instantanée pour retrouver un contenu spécifique.



Dans ses applications, BIREME utilise des versions spécifiques et propriétaires de CISIS, avec des clés courtes jusqu'à 16 caractères, ainsi que des clés longues de 17 à 60 caractères, parce que dans le secteur des sciences de la santé il est courant d'utiliser des termes plus longs.

Maintenance de l'information dans la base de données

Pour se transformer en un réel outil de manipulation de données ou de transfert de connaissances, une base de données doit permettre l'ajout, la suppression et la modification du contenu.

Les bases de données CDS/ISIS mettent en oeuvre différents mécanismes pour ces opérations de base que l'on va décrire maintenant.

Les enregistrements que l'on ajoute à la base de données, comme l'illustre la figure 3, sont intégrés au fichier de données de la base (MST) et sont référencés dans le fichier des pointeurs (XRF).

Les enregistrements que l'on supprime de la base de données sont marqués comme non valables ou sont logiquement effacés, en attendant d'être physiquement effacés (élimination des données du fichier de la base de données) lors d'une opération de maintenance de la base de données, comme nous le verrons ensuite.

Les enregistrements que l'on modifie – étant donné que leur taille peut augmenter – suivent le principe vu plus haut pour l'ajout d'enregistrements et les règles d'occupation de l'espace de stockage ; les données d'origine restent à leur emplacement de départ mais sont marquées comme non valables. Ainsi, à cause de chaque modification d'enregistrement, l'espace occupé par la base de données augmente ; c'est en effet le prix de la flexibilité des enregistrements puisqu'ils sont composés de champs de taille variable.

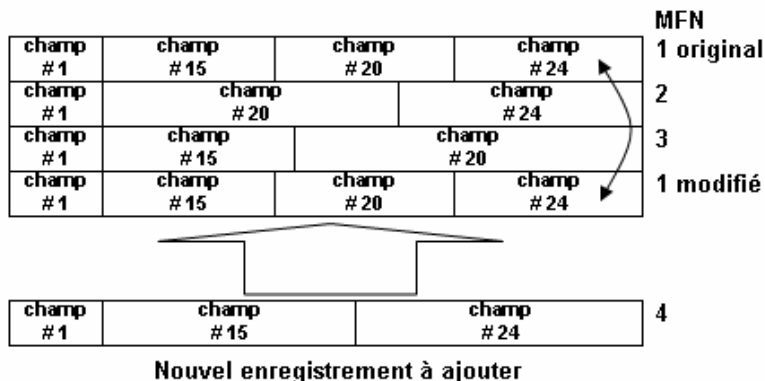


Figure 3 : Base de données en cours d'utilisation

Pour éviter cet effet indésirable, les opérations de maintenance de la base de données doivent prévoir la réorganisation régulière de la base, afin d'en restructurer le contenu. Une telle réorganisation peut être menée à bien de différentes manières. La plus basique de toutes consiste à balayer le fichier maître de la base de données enregistrement par enregistrement en suivant les MFNs, et à créer un nouveau fichier maître, dans lequel les registres sont intégrés dans leur ordre de numérotation. On élimine ultérieurement le fichier maître balayé et on renomme le fichier maître créé pour qu'il remplace l'ancien. Une autre façon d'effectuer cette réorganisation, beaucoup plus courante pour les utilisateurs du CDS/ISIS pour Windows, est d'exporter la base de données au format ISO, et ensuite de l'importer dans la base originale.

De cette manière, on redistribue les données de la base représentée dans la figure 3. Le résultat est présenté ci-dessous à la figure 4 :

champ # 1	champ # 15	champ # 20	champ # 24	MFN 1
champ # 1	champ # 20		champ # 24	2
champ # 1	champ # 15	champ # 20		3
champ # 1	champ # 15		champ # 24	4

Figure 4 : Base de données réorganisée

En général, après la réorganisation on effectue une ré-indexation complète de la base de données, pour s'assurer que les enregistrements effacés sont bien éliminés des anciens index du fichier inversé.

Jeux de caractères utilisés

En fonction de l'environnement, on utilise un jeu de caractères approprié pour représenter les caractères imprimables. Par exemple, lorsqu'on travaille sous MS-DOS (Microsoft Disk Operating System) on se sert du jeu de caractères ASCII, ou lorsqu'on travaille sous Windows on se sert du jeu de caractères ANSI.

Par exemple, le code 162 (0xA2 en hexadécimal) du jeu ASCII, représente le caractère en minuscules avec un accent aigu (ó), tandis que dans le jeu ANSI il représente le symbole des centimes (¢).

De même, en fonction de la localisation des ordinateurs (c'est-à-dire le pays et les paramètres régionaux), il peut y avoir des variations dans les jeux de caractères. Dans le cadre du codage ASCII, le jeu de base est celui que l'on appelle CODE PAGE (page de codes) 437 qui répond aux besoins des caractères imprimables pour les Etats-Unis ; on peut utiliser au contraire la page de codes 850 qui répond aux besoins des caractères imprimables des pays qui utilisent des langues latines, surtout l'espagnol et le portugais.

Par exemple, le code 199 (0xC7 en hexadécimal) de l'ASCII CP 437 est un caractère graphique pour dessiner les bordures, tandis qu'avec l'ASCII CP 850 cela correspond au caractère en minuscules avec tilde (ã).

On y supplée grâce à une méthode de conversion des codes de caractères imprimables entre les différents jeux de caractères, dans le cadre de CDS/ISIS ; cette méthode peut s'appuyer sur l'emploi de bases de données spéciales, appelées GIZMO. Une base gizmo comporte deux champs de données et met en rapport le premier champ avec les données d'entrée ; il donne en sortie la donnée équivalente qui est contenue dans le second champ de l'enregistrement. La commande qui rend possible cette conversion sera présentée plus tard.

Il y a quatre bases gizmo disponibles pour :

- lire les caractères ANSI et les convertir en ASCII CP 437 (gans437) ;
- lire les caractères ANSI et les convertir en ASCII CP 850 (gans650) ;
- lire les caractères ASCII CP 437 et les convertir en ANSI (g437ans) .;
- et en dernier lieu, lire les caractères ASCII CP 850 et les convertir en ANSI (g850ans).

Il faut savoir qu'un système de saisie de données Web emploie l'ensemble de caractères ISO-8859-1 qui a les mêmes codes et signes imprimables que le jeu de caractères ANSI ; par conséquent, il s'agit d'une base de données alimentée dans un système Web qui, lors d'un transfert sous DOS, doit convertir ses caractères avec le codage approprié.

Différences entre plate-formes

Les modifications de système d'exploitation peuvent entraîner un changement de plate-forme, par exemple, de Windows à Linux, ou à Unix, etc. Comme il y a des caractéristiques spécifiques dans l'enregistrement des fichiers dans chacune de ces plate-formes, le changement de plate-forme peut provoquer des erreurs dans la lecture du fichier maître et du fichier inversé de CDS/ISIS.

Pour éviter ce type de difficulté, le *package* CISIS dispose de deux applications qui effectuent les modifications nécessaires dans les fichiers pour - actuellement - neuf plate-formes, à savoir : Linux, HP-UX, Sun, Alpha, Vax, Unisys, Mpe, CDC et Windows. L'une des applications sert pour la conversion du fichier maître, l'autre pour la conversion du fichier inversé.

C'est pourquoi, avant le transfert d'une base de données vers un système d'exploitation différent, il faut vérifier la nécessité de changer l'ensemble de caractères et le format d'enregistrement des fichiers de la base de données (et le fichier inversé, si nécessaire) pour pouvoir mener à bien le processus en trois étapes : conversion de l'ensemble de caractères, conversion du format du fichier et transfert vers l'autre système, ainsi que l'illustre la figure 5 :

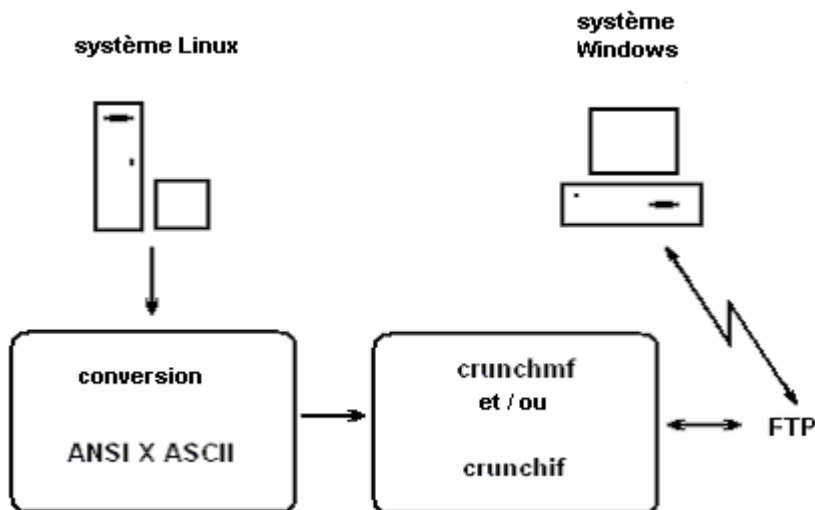


Figure 5 : Transfert de bases de données entre systèmes d'exploitation

Initiation à l'utilisation de CISIS

La version actuelle de CISIS est la 4.3 (année 2004), qui comporte diverses améliorations, caractéristiques et possibilités et qui se compose de différentes 'commandes', comme par exemple : mx, crunchmf, crunchif, msrt, mxtb, mxcp, mkxrf et bien d'autres.

MX l'utilitaire de CISIS

La commande d'usage courant de CISIS pour effectuer lecture, écriture, restauration et inversion des bases de données est MX.



En tapant mx<entrée> dans la ligne de commande, on obtient un résumé des paramètres disponibles pour MX. En réalité, tous les composants de CISIS donnent ce type d'informations pour les paramètres de base, lorsqu'ils sont appelés depuis une ligne de commande, sans aucune mention de paramètre.

Il suffit de donner le nom d'une base de données pour que mx procède à sa lecture, enregistrement par enregistrement, en commençant par le premier. Ainsi la commande **mx lil** retourne :

```

mx lil
mfn= 1 [DELETED]
..
mfn= 2 [DELETED]
..
mfn= 3 [DELETED]
..
mfn= 4 [DELETED]
..
mfn= 5 [DELETED]
..
mfn= 6
1 «BR1.1»
2 «000006»
3 «BR1.1/7.02»
4 «LILACS»
5 «MC»
6 «am»
10 «Banta, David»
12 «The uses of modern technologies: problems and perspectives for
industrialized and developing countries»
14 «29-46»
17 «Conselho Nacional de Desenvolvimento Científico e Tecnológico, ed»
17 «Organización Panamericana de la Salud, ed»
18 «Conferência Interamericana sobre a Avaliação Tecnológica em Saúde»
20 «234»
40 «En»
52 «Brasil. Ministério da Saúde»
52 «Brasil. Ministério da Educação»
52 «Brasil. Ministério da Previdência e Assistência Social»
52 «Instituto Nacional de Assistência Médica e Previdência Social»
52 «Instituto de Pesquisas Econômicas e Sociais»
52 «Organización Panamericana de la Salud»
52 «Conselho Nacional de Desenvolvimento Científico e Tecnológico»
53 «Conferência Interamericana sobre a Avaliação Tecnológica em Saúde»
54 «14-18 nov. 1983»
55 «19831115»
56 «Brasília»
57 «BR»
62 «Conselho Nacional de Desenvolvimento Científico e Tecnológico»
64 «1985»
65 «19850000»

```

Il faut préciser qu'à chaque enregistrement affiché la commande s'arrête et attend que l'utilisateur demande un nouvel enregistrement avec la touche <ENTREE>. Cette invite de commande (ou prompt) est signalée par la présence de deux points accolés (.), sur la console.

Si l'on souhaite vérifier un enregistrement précis dont on connaît le numéro, la commande accepte la syntaxe suivante : *mx <base> from=<reg_num>*, par exemple :

```

D:\Documentos\teste>mx lil from=10000
mfn= 10000

```

```

1 «BR1.1»
2 «010000»
4 «IMLA»
4 «LILACS»
5 «S»
6 «as»
10 «Manterola, A»
12 «Coma y alteraciones de conciencia en el niño.»
13 «Coma and consciousness changes in children»
14 «45-58»
30 «Pediatría (Santiago de Chile)»
31 «25»
32 «1/2»
40 «Es»
64 «1982»
65 «19820000»
76 «INFANTE»
76 «NINO»
76 «HUMAN»
87 «^dCOMA»
87 «^dCONSCIOUSNESS DISORDERS»
90 «b»
..

```

Si l'on ne souhaite pas s'arrêter à chaque enregistrement, on peut inclure le paramètre *now* dans la commande *mx*, ce qui donne une lecture séquentielle de tous les enregistrements de la base, à partir d'un certain point, s'il est précisé, ou jusqu'à la fin.

Il est possible d'utiliser et de combiner différents paramètres pour utiliser *mx*, ce qui lui confère une très grande souplesse. Pour accroître cette souplesse, on peut associer des commandes natives du système d'exploitation, comme par exemple, le comptage des lignes, un tri avec l'élimination d'éléments répétitifs, etc.

Il est important de souligner que, pour effectuer le changement de caractères, le paramètre *gizmo* est disponible, ce qui permet d'utiliser les bases *gizmo* pour changer l'ensemble des caractères d'une base de données, par exemple :

```
mx lilacs gizmo=g850ans create=lilans -all now
```

Ici, à partir d'une base Lilacs utilisant des caractères ASCII, on crée une autre base avec des caractères ANSI.

Autres commandes de CISIS

crunchmf et crunchif

Les commandes de conversion de format de fichier `crunchmf` y `crunchif` s'utilisent lorsque l'on souhaite migrer une base de données et/ou un fichier inversé vers une autre plate-forme qui n'est pas celle d'origine, par exemple, migrer une base Windows vers HP-UX.

La manière d'utiliser les deux commandes est identique et nécessite qu'on précise le fichier à lire (fichier maître ou fichier inversé), le nom du fichier à créer dans le nouveau format d'enregistrement (fichier maître ou fichier inversé) et le système d'exploitation retenu, pour ne citer que les possibilités les plus fréquemment utilisées.

Ci-dessous un exemple avec la base DeCS, actuellement sur une plate-forme Linux et que l'on veut transférer vers une plate-forme Windows.

```
crunchmf decs win/decs target=pc tell=5000
```

msrt

Une commande assez utile lors de la réalisation d'un rapport est `msrt`, qui effectue l'organisation des enregistrements d'un fichier maître en fonction d'un critère basé sur le contenu des champs de la base de données. Ainsi il est possible de classer la base par date de publication (à supposer qu'un tel champ existe dans la base), ou en fonction d'un critère plus complexe, comme par exemple par ordre alphabétique d'auteurs et de titres.

Ci-dessous un exemple avec la base Lilacs que l'on classe par date de publication du document (champ 65), avec un critère de tri sur le titre du document.

```
msrt lilacs 160 "s(v65,v18,v12)"
```

mxtb

Une autre commande d'une grande utilité est `mxtb`. Sa fonction est de mettre les données de la base sous forme de tableau en créant une nouvelle base (dont le nom

est donné dans l'appel de la commande) ; cette base a une structure standardisée qui indique la fréquence des données, extraites conformément au format indiqué. Les données à mettre sous forme de tableau peuvent être sélectionnées à partir d'une équation de recherche dans la base de données, ce qui permet de limiter les données à mettre sous forme de tableau.

Ci-dessous un exemple dans lequel on met sous forme de tableau les années de publication (les quatre premiers caractères de v65) d'un titre de publication en particulier, dans la base Lilacs.

```
mxtb lilacs create=rbopd "4:v65.4" "bool=Rev. Bras. Oftalmol"
```

Dans le tableau 1, sont présentés les différents paramètres d'appel de manière à former une liste de situations ordinaires pour servir de guide dans l'exécution de tâches courantes.

Tâche	Syntaxe
Lire à partir d'un enregistrement	<i>mx <base> from=<mfn_départ></i>
Lire une série d'enregistrements	<i>mx <base> from=<mfn_initial> to=<mfn_final></i>
Fixer le nombre d'enregistrements	<i>mx <base> +control count=-0</i>
Créer une copie de la base	<i>mx <base_in> create=<base_out> -all now</i>
Créer une copie de la base sans les enregistrements effacés	<i>del <base_out.mst> (ou rm <base_out.mst>) del <base_out.xrf> (ou rm <base_out.xrf>) mx <base_in> append=<base_out> -all now</i>
Générer un fichier ISO de la base	<i>mx <base> iso=<fichier.iso> -all now</i>
Créer une base à partir d'un fichier ISO	<i>mx iso=<fichier.iso> create=<base> -all now</i>
Créer une base à partir d'un fichier ISO en stockant le mfn dans un champ particulier	<i>mx iso=<fichier.iso> create=<base> "proc='vn" -all now</i>
Inverser une base de données	<i>mx <base> "fst=<clé_de_tri>" fullinv=<fichier_inversé></i>
Inverser une base de données avec un fichier FST spécifique	<i>mx <base> "fst=<@fichier.fst>" fullinv=<fichier_inversé></i>
Inverser une base utilisant des caractères ANSI	<i>mx <base> "fst=<@fichier.fst>" fullinv/ansi=<fichier inversé></i>
Convertir le jeu de caractères de la base	<i>mx <base> gizmo=<base_gizmo> create=<base_out> -all now</i>
Convertir la base pour une autre plate-forme	<i>crunchmf <base> <base_out> target=<plate-forme></i>

Tâche	Syntaxe
Convertir le fichier inversé pour une autre plate-forme	<i>crunchif <fichier_inversé> fichier_inversé_out target=<plate-forme></i>
Ecrire un fichier lisible par Excel	<i>mx <base> "pft= "/vn "/, ,"/vn "/, ,"/vn "//" -all now > <fichier.csv></i>
Importer un fichier Excel (CSV)	<i>mx "seq=<fichier.csv>," create=<base_out> -all now</i>
Extraire les données d'une base en les triant	<i>mx <base> "pft=<format>" -all now sort -u (linux)</i>
Effectuer des statistiques sur les champs, les sous-champs et les caractères de la base de données	<i>mxfo <base> create=<base_out> ou noedit</i>
Trier une base en fonction d'un champ	<i>msrt <base> <#caract_clé> <format_extraction></i>
Mettre sous forme de tableau le contenu d'un champ de la base	<i>mxtb <base> create=<base_out> <longueur:format> [class=nnnnn]</i>

Tableau 1 : Commandes les plus courantes avec CISIS

Aspects pratiques

Equivalence entre les commandes Linux et Windows

La plupart des fonctions utilisées dans un système d'exploitation donné ont leur équivalent dans les autres. Ci-dessous, dans le tableau 2, une liste des équivalences entre Linux, DOS et Windows.

LINUX	DOS	WINDOWS
<i>clear</i>	<i>cls</i>	Ne s'applique pas
<i>ls</i>	<i>dir /w</i>	Volet de visualisation dans l'explorateur Windows (à droite)
<i>ls -l</i>	<i>dir</i>	Volet de visualisation dans l'explorateur Windows (à droite)
<i>mkdir</i>	<i>mkdir</i>	Crée un nouveau répertoire
<i>rmdir</i>	<i>rmdir</i>	Efface un répertoire vide
<i>cd</i>	<i>cd</i>	Navigue dans l'explorateur Windows vers un autre dossier
<i>rm</i>	<i>del</i>	Efface un (ou plusieurs) fichier(s)
<i>rm -r</i>	<i>deltree</i>	Efface un répertoire (et son contenu)
<i>ftp</i>	<i>ftp</i>	Démarre une session FTP(ws-ftp)

Tableau 2 : Equivalences des commandes entre les systèmes d'exploitation

Commandes FTP les plus courantes

Quand il faut transférer des fichiers par FTP (File Transfer Protocol) et lorsque l'application fonctionne en mode console et non avec une interface graphique, il existe une liste de commandes communes à la plupart des programmes ftp. Dans le tableau 3 on trouve la liste de ces commandes.

Commande	Explication
<i>open</i>	Permet de se connecter à un serveur FTP
<i>ls</i>	Liste tous les fichiers du répertoire courant
<i>cd</i>	Change de répertoire courant
<i>bin</i>	Active le mode de transfert binaire
<i>asc</i>	Active le mode de transfert ASCII
<i>put</i>	Envoie un fichier vers le serveur
<i>mput</i>	Envoie plusieurs fichiers vers le serveur
<i>get</i>	Récupère un fichier sur le serveur
<i>mget</i>	Récupère plusieurs fichiers sur le serveur
<i>prompt</i>	Active/désactive le mode interactif

Tableau 3 : Commandes FTP les plus courantes

Conversion du jeu de caractères

Pour effectuer la conversion du jeu de caractères de la base de données, on utilise la commande **mx** qui est le composant le plus puissant et le plus utilisé de CISIS, avec une base gizmo, conformément à ce qui a été indiqué plus haut.

En supposant que l'on veuille changer le jeu de caractères ASCII Code Page 850 en ANSI, on utilise la commande suivante :

```
mx BASE_IN gizmo=g850ans create=BASE_OUT now -all
```

La base de données de départ utilisée dans cet exemple est appelée BASE_IN, mais elle peut avoir évidemment un autre nom. La base gizmo, dans cet exemple, est dans le répertoire courant et une autre base de données appelée BASE_OUT (elle peut également avoir un autre nom) sera créée avec le jeu de caractères souhaité.

Dans l'exemple ci-dessous, la base LILACS, utilisant le jeu de caractères ANSI, est convertie en créant la base LIL850, au moyen de la base gizmo ANSI → ASCII CP850 qui se trouve dans le répertoire **tabs**, voisin ou courant.

```
mx lilacs gizmo=../tabs/gans850 create=lil850 now -all
```

Transfert de bases entre systèmes d'exploitation

Si l'on souhaite transférer une base de données (seulement le fichier maître) d'une plate-forme à une autre, avec des systèmes d'exploitation différents, il faut utiliser la commande **crunchmf**, qui est l'un des composants de CISIS.

Supposons de que la base soit sur un serveur Linux et que l'on désire la transférer vers une machine Windows, on utilise la commande suivante :

```
crunchmf BASE_ORI BASE_DEST target=pc
```

La base de données de départ utilisée dans cet exemple est appelée `BASE_ORI`, mais elle peut avoir évidemment un autre nom. On créera une base de données de destination appelée `BASE_DEST` (elle peut également avoir un autre nom) ; la conversion du format physique est désignée par le paramètre **target=** qui est dans cet exemple **pc**, c'est-à-dire un système qui utilise Windows.

Dans l'exemple ci-dessous, la base GANSNA, actuellement sur un PC avec Windows, sera convertie pour être transférée vers une machine Linux, dans le répertoire LNX.

```
crunchmf gansna lnx\gansna target=linux
```

Le même principe est appliqué pour un fichier inversé, mais en utilisant la commande **crunchif**, autre composant de CISIS.

Références bibliographiques

1. UNESCO. *Mini-micro CDS/ISIS: Reference manual (version 2.3)*. Organized by Giampaolo Del Bigio. Paris: United Nations Educational, Scientific and Cultural Organization, 1989. 286 p. ISBN 92-3-102-605-5.
2. BUXTON, Andrew, HOPKINSON, Alan. *The CDS/ISIS for Windows Handbook* [online]. Paris: United Nations Educational, Scientific and Cultural Organization, 2001 [cited 30 August 2006]. 164 p. Available from internet: <<http://bvsmodelo.bvs.br/download/winisis/winisis-handbook-en.pdf>>.
3. SUTER, Tito. "Prehistoria" e historia del MicroISIS [online]. In: *Manual para instructores de Winisis*. Buenos Aires: Centro Atómico Constituyentes (CAC), Comisión Nacional de Energía Atómica (CNEA), 1999 [citado el 30 Agosto 2006]. p. 21-26. Disponible en internet: <<http://www.cnea.gov.ar/cac/ci/isis/isidams.htm>>.

Glossaire

- **Base de données.** Collection de données structurée pour qu'il soit possible d'y accéder et de les manipuler facilement. Est formée d'unités d'information appelées enregistrements dont les différents éléments sont représentés par des champs et des sous-champs. Par exemple, dans un "fichier clients", chaque client représente un enregistrement possédant plusieurs champs, comme "NOM", "CODE DU CLIENT", "TÉLÉPHONE" etc.
- **Bases de données bibliographiques.** Version électronique d'un catalogue ou d'un index bibliographique.
- **CDS/ISIS - MicroISIS.** Logiciel développé et maintenu par l'UNESCO pour le traitement des données bibliographiques.
- **Champ.** Élément d'un enregistrement permettant de stocker une information spécifique. Voir *Base de données*.
- **Clé.** Expression identifiant une ou plusieurs informations d'un type déterminé et pouvant être utilisée en recherche.

- **Enregistrement.** Ensemble structuré de données qui permet de stocker tous les éléments de données d'une unité d'information. Voir Base de données.
- **Fichier.** En informatique, un ensemble de données qui peut être enregistré dans un certain dispositif de stockage. Les fichiers de données sont créés par des applications, comme par exemple un traitement de textes.
- **Format électronique.** Toute forme de stockage, récupération et présentation d'information susceptible de faire l'objet d'une transmission en ligne ou d'un enregistrement sur des supports magnétiques ou optiques.
- **Format ISO (échange de données).** Format standard défini par l'ISO pour l'échange de données entre institutions, réseaux et utilisateurs. Se réfère à la norme ISO 2709.
- **Format LILACS.** Format de description bibliographique défini par BIREME, basé sur le Reference Manual for Machine-readable Bibliographic Descriptions de l'UNISIST.
- **Indexation.** Procédure pour identifier et décrire le contenu d'un document avec des termes représentant les sujets correspondant à ce document, dans le but de le retrouver ultérieurement.
- **Sauvegarde.** Procédure par laquelle un ou plusieurs fichiers sont dupliqués sur un autre dispositif de stockage (bande ou disque), pour produire une copie de sécurité qui pourrait être restaurée si certaines données sont accidentellement effacées ou s'il se produit des dommages physiques affectant les données originales.

- **Sous-champ.** Élément qui contient la plus petite partie de l'information dans un champ, et qui peut n'avoir aucune signification si elle est considérée de manière isolée.
- **UNISIST.** Programme intergouvernemental relatif à la coopération dans le domaine de l'information scientifique et technologique.