



12th ANNUAL MEETING OF THE
DOBIS/LIBIS USERS GROUP

**Téléchargement et reformatage de données DOBIS/LIBIS
pour la réutilisation dans un système local de gestion de fichiers**

Francesco Dell'Orso, Adriana Panza

Univ. of Perugia-Univ. of Modena, Italy



Coimbra - Portugal

14 - 17 September 1993

**Téléchargement et reformatage de données Dobis/Libis
pour la réutilisation dans un système local de gestion de fichiers**

Francesco Dell'Orso
(Université de Pérouse, Italie)

Adriana Panza
(Université de Modène, Italie)

University of Coimbra
Coimbra - Portugal
14-17 September 1993

Francesco Dell'Orso
(*Université de Pérouse, Italie*)
Adriana Panza
(*Université de Modène, Italie*)

Traduction: Odile Martinez (Pérouse)

Manuscrit terminé le 11 septembre 1993

SOMMAIRE

1	Introduction	1
2	Téléchargement: ressources, demande, moyens	1
3	Les phases et les opérations du téléchargement dans le détail	3
4	Téléchargement des données bibliographiques en milieu Dobis: expériences documentées	4
5	Notre intervention	4
5.1	<i>Objectifs et orientations</i>	4
5.2	<i>Objectifs à négliger</i>	5
5.3	<i>Logiciels utilisés</i>	6
5.3.1	Logiciels d'émulation, de capture des données, de traitement de texte	6
5.3.2	Logiciels de reformatage et de conversion	6
5.3.3	Logiciels de gestion de fichiers	6
5.4	<i>Déroulement de l'expérience</i>	7
5.4.1	Approche sans modification de Dobis	7
5.4.2	Approche avec modification de Dobis	8
5.5	<i>Recherche, repérage et capture des données (phases 1 et 2)</i>	10
5.6	<i>Reformatage du texte, premier traitement des données (phase 3a)</i>	10
5.7	<i>Conversion, deuxième traitement (phase 3b)</i>	12
5.8	Résultats obtenus et appréciations	13
5.8.1	Résultats obtenus avec tous les logiciels testés	13
5.8.2	Résultats obtenus uniquement avec quelques logiciels	14
5.8.3	PBS/Pro-Cite	14
5.8.4	Limites relevées dans tous les logiciels de conversion	14
5.8.5	LBSPC/Libis-Net (Leuven)	15
6	Propositions	15
6.1	<i>Propositions à court terme</i>	15
6.2	<i>Propositions pour un développement plus consistant</i>	16
7	Limites et problèmes	17
8	Annexe 1: produits et producteurs	19

LISTE DES ILLUSTRATIONS

Fig. 1	DOBIS Catalogage - Description complète: notice sur deux écrans	9
Fig. 2	DOBIS Catalogage - Description complète: après édition par traitement de texte	11
Fig. 3	OPAC LIBI - Description complète modifiée	11
Fig. 4	Synthèse des résultats du reformatage	12
Fig. 5	Conversion: Fangorn, indication des caractéristiques des données en entrée	12
Fig. 6	Conversion, Data Magician: caractéristiques du fichier d'entrée	13

1 Introduction

Le travail présenté dans cet article a été réalisé en collaboration par des bibliothécaires et des informaticiens¹, et il s'adresse à des bibliothécaires et à des informaticiens, mais il privilégie le point de vue de l'utilisateur final, bibliothécaire, lecteur ou chercheur. La solution que l'on a cherché à donner au problème soulevé est telle qu'elle peut être directement appliquée par les usagers, et ultérieurement adaptée par eux.

Le problème traité est celui du transfert de données à partir de fichiers centraux vers les fichiers locaux d'un micro-ordinateur afin de les réutiliser, ce que la littérature définit en termes un peu généraux comme télédéchargement (*downloading*). Cette opération comprend en effet aussi le reformatage, après le télédéchargement, et le téléchargement de données dans un autre système receveur pour leur réutilisation.

On entend généralement par télédéchargement le prélèvement, l'acheminement ou le transfert, c'est-à-dire la copie de données d'un ordinateur à un autre, ou d'un fichier à un autre sur le même ordinateur; l'opération inclut le transfert de fichier (*file transfer*), et pour certains elle pourrait s'y ramener entièrement². Ce qui nous intéresse ici, c'est justement le mouvement du plus grand vers le plus petit, et par surcroît la destination finale: le but est en effet d'ajouter au plus petit les données déchargées du plus grand, c'est-à-dire d'aboutir à une réutilisation des données (*up-loading*).

Nous considérons évidemment la question dans le contexte de Dobis, car il nous semble que d'autres systèmes de gestion des bibliothèques l'ont fait ou sont en train de le faire, tandis que jusqu'à présent Dobis l'a seulement subi ou réalisé à partir de prototypes, paradoxalement trop avancés pour les utilisateurs³. Etant donné que l'on parle du successeur de Dobis comme d'un événement assez proche, et si l'on reconnaît une certaine valeur au télédéchargement, le moment est opportun pour mûrir les intentions et les décisions à ce sujet.

Synthétiquement, les phases de ce que l'on appelle télédéchargement pourraient être au nombre de deux⁴:

- 1 recherche & capture
- 2 conversion & importation/téléchargement

2 Télédéchargement: ressources, demande, moyens

Il semble bien que depuis environ une quinzaine d'années les fichiers des services bibliothécaires et de documentation ont été informatisés. Parmi les différents types de fichiers, on peut relever:

- 1 catalogue bibliographique: références secondaires qui décrivent des documents primaires (*reference database*). Le catalogue principal comprend souvent:
 - des données codées concernant l'identité des publications: pays, langue de publication, prix, date de parution, etc.
 - des données sur les exemplaires: état des collections, nombre de copies, localisation, état de conservation, etc.
- 2 fichiers de données administratives concernant les documents (*source database*): date de commande, date d'entrée, prix payé, etc.
- 3 fichiers-répertoires (*referral database*) sur:
 - a) les bibliothèques: nom, adresse, fonds, horaires, services, etc.
 - b) les utilisateurs: nom, profession, âge, adresse, etc.
 - c) les fournisseurs: nom, catégorie, pays, etc.

En tant que personnel, nous introduisons et maintenons les données, nous les visualisons, nous les imprimons. Les données disponibles ont augmenté. L'utilisation du système de la part des opérateurs et des usagers s'est accrue. Les

¹ Adriana Panza travaille comme informaticienne à l'Université de Modène, Francesco Dell'Orso travaille comme bibliothécaire à l'Université de Pérouse.

² Transférer, exporter, prélever, capturer, sont des termes qui maintiennent une ambiguïté quant au fait que l'original semble disparaître, alors qu'il demeure. Dans tous les cas ces opérations aboutissent à une copie identique à l'original, elles répètent ailleurs une image des données figurant dans la source. Mais étant donné qu'elles sont 'télé', ce n'est pas un simple transfert 'de/à' sur un pied d'égalité, mais une 'copie' du 'plus' (source) vers le 'moins' (cible). Il faut en effet relever comme différences: la quantité de données figurant dans le système serveur, sa puissance, la distance entre les deux contextes. Si l'on excepte la différence entre ces plans, le télédéchargement n'est qu'un transfert de données.

³ Cf. la note 12 p. 4.

⁴ Certains auteurs soutiennent, et c'est aussi notre avis - cf. § 3 p. 3 -, qu'il s'agit de deux passages aussi bien dans la première que dans la deuxième phase, et donc de 4 en tout. En ce qui concerne la première phase, le fichier de sortie en ASCII peut être produit également par le système hôte, mais pour le transférer il faut en sortir et recourir à l'émulateur, et, peut-être, à d'autres utilitaires. Quant à la seconde phase, dans le cas des logiciels plus ou moins généraux de conversion, elle est nettement séparée en: conversion dans un format et téléchargement dans un système, et là il n'y a pas de correspondance de 1 à 1 entre format et système. Cette phase est également articulée en deux moments dans le cas de logiciels qui incorporent la fonction de conversion, parce qu'il y a toujours une phase, distincte, de préparation des tables de transcodification.

fichiers sur papier devraient théoriquement avoir diminué, c'est du moins ce que l'on dit. Même les utilisateurs recourant à l'OPAC (catalogue public en ligne) visualisent et utilisent les données.

Il est donc acquis, et compréhensible, que les exigences et les demandes des utilisateurs augmentent. Les demandes ne concernent pas habituellement les données de toute la base, mais uniquement certains de ses sous-ensembles. Les utilisateurs finals de fichiers bibliographiques peuvent demander à bénéficier de l'élaboration des données pour les objectifs suivants:

- citations à inclure dans une communication, un article, un programme de recherche, un curriculum
- présenter les résultats d'une recherche dans une base de données
- préparer des statistiques (prix, date, code du pays, langue, type de fournisseur...)
- imprimer des produits particuliers par leur structure et leur présentation: catalogues dotés de différents index, utilisation de caractères raffinés
- effectuer des échanges avec d'autres utilisateurs de systèmes semblables ou différents
- réutilisation personnelle (chercheur, bibliothèque, centre, DSP...) pour bases de données personnelles ou locales, ou de toute façon autres que celles du système central
- traitement supplémentaire en différé⁶.

Qui satisfait normalement ces demandes? Le Centre de calcul (CDC).

Le CDC peut répondre de différentes façons: en effectuant l'élaboration du début jusqu'à la fin, en cédant le résultat de l'élaboration centrale qui sera traité ultérieurement, ou bien en ne fournissant que les données brutes qui seront ensuite manipulées par l'utilisateur.

Qu'entraîne cette solution?

- beaucoup de travail pour le CDC: le personnel a ainsi moins de temps pour s'occuper des recherches, des améliorations, des innovations à appliquer au système central
- le plus souvent le CDC fournit des produits standard à partir de ceux qui sont disponibles dans l'ordinateur central. Pour excellents que soient ces produits, la nature humaine, et l'augmentation constante de solutions informatiques variées et puissantes ont donné force de postulat à la conviction qu'aucun standard ne pourra satisfaire toutes les exigences particulières des usagers
- entre les utilisateurs qui demandent, et le CDC qui distribue, s'instaure le même rapport qu'entre structures décentralisées et centre de services: tous deux vivent la demande comme une obligation, mais c'est en définitive le Centre qui a le pouvoir de la satisfaire
- d'autre part, si c'est le CDC qui se charge de l'élaboration principale, bibliothécaires et utilisateurs sont dégagés de l'obligation d'acquérir des compétences, d'effectuer un travail, d'engager aussi leur responsabilité dans le secteur informatique, au sens large.

La gestion centralisée d'un système centralisé pour toutes les formes d'exploitation des ressources qui y sont accumulées offre-t-elle la bonne solution? Oui, dans certains cas: on ne peut pas tout décentraliser. Mais elle n'est pas la seule solution. Elle l'est même de moins en moins.

Aujourd'hui, et ça ne date pas d'hier, le micro-ordinateur - *PC Personal Computer* - offre matériels et logiciels ad hoc ayant des coûts, des prestations, un potentiel, une souplesse tels qu'ils sont en mesure de satisfaire les exigences individuelles.

Le téléchargement et la réutilisation des données ont donc une justification parce qu'on a: d'une part l'offre de ressources, les données (catalogues en ligne de bibliothèques (*Online library catalogs*), bases de données en ligne ou sur CD-ROM), d'autre part la demande des utilisateurs et les moyens pour la satisfaire de bien des façons.

Enfin, peut-on encore considérer comme utilisateurs d'un système bibliothécaire uniquement ceux qui le fréquentent effectivement, et ceux qui, pour des raisons institutionnelles, peuvent utiliser le travail du CDC local? Le développement des réseaux de communication et des logiciels permettant leur utilisation a permis d'accéder à des centaines de catalogues de bibliothèques et de fichiers documentaires. Dans cette bibliothèque virtuelle, sont virtuels non seulement les documents que l'on peut repérer, emprunter ou obtenir en copie partielle, mais aussi les utilisateurs du monde entier qui peuvent utiliser notre catalogue.

Interdire ou entraver le téléchargement veut dire:

- ne pas concevoir le système central comme un *serveur* de ressources
- faire converger toutes les demandes d'élaboration vers le CDC

⁵ Cf. Miido, Helis. "Combined use of internal databases, CD-Roms and online databases: problems and solutions." *Online Information 89: 13th International Online Information Meeting, London 12-14 December 1989*, 349-362. Oxford and New Jersey: Learned Information, 1989.

⁶ Cela a un sens en particulier pour le téléchargement à partir d'une base de données en ligne: les coûts seraient en effet différents. Cf. Lardy, Jean-Pierre, and Pascal Bador. "Les techniques de téléchargement et de télétransmission de quelques logiciels de communication." *Documentaliste* 27, no. 2 (1990): 63-68, e Benharrat, Alia. "Pratique du télétransmission: petit guide technique." *Documentaliste* 28, no. 4-5 (1991): 214-217. Quant aux catalogues de bibliothèques en ligne, imprimer en différé permet de ne pas engorger le système.

- obliger l'utilisateur final, désirant créer des fichiers dérivés, à saisir lui-même toutes les données au clavier. Métaphoriquement, ce serait comme indiquer, au lieu du pont entre deux hauteurs, le gué situé au fond de la vallée.

On pourrait voir dans ces considérations l'expression d'un état d'alerte et d'une urgence, pourtant le téléchargement et la réutilisation des données ne sont pas du tout une nouveauté des années quatre-vingt dix.

En 1982, au cours de la Online Conference d'Atlanta, Carlos Cuadra, Président de la Cuadra Associates, "sollicitait tous les distributeurs de base de données à décider rapidement d'une politique concernant le téléchargement des données et leur stockage dans des micro-ordinateurs locaux"⁷.

Toutefois, en dehors du monde anglo-saxon, le problème a eu une résonance très réduite, et continue à en avoir moins que son pendant: le téléchargement (*uploading*)⁸. En 1991, un article élaboré dans le contexte de la New York University Law Library pouvait encore commencer par ces mots: "Pourquoi télécharger?" ("Why download?")⁹. Si on se demande encore 'pourquoi', c'est parce qu'on peut aussi ne pas être en faveur du téléchargement. Nous reviendrons sur cette question en conclusion.

3 Les phases et les opérations du téléchargement dans le détail

Nous avons déjà dit que la procédure globale comprenant le repérage, la capture et la réutilisation de données d'une ou de plusieurs bases de données centrales vers des bases de données habituellement plus petites est appelée téléchargement. En réalité, les différentes phases sont variées, et le téléchargement n'est qu'une d'entre elles. On peut distinguer au moins les phases suivantes:

- 1 interrogation et repérage
- 2 capture et copie
- 3 manipulation et conversion
- 4 chargement et réutilisation

1 Interrogation et repérage

Démarrage du micro-ordinateur, lancement des logiciels résidents en mémoire TSR (*terminate and stay resident*), du logiciel d'émulation d'un terminal vidéo, VDU (*video terminal unit*). Accès à l'hôte et à l'application. Opération de recherche et de repérage des données.

2 Capture et copie, déchargement des données

Prélèvement des données un écran après l'autre (*screen-by-screen*), capture d'une session d'interrogation (*log file*), ou bien capture d'une série de références formant un fichier. Cela en interceptant les données, soit sur écran, soit grâce à une commande d'enregistrement direct du fichier sur disque, ou après qu'elles ont été dirigées vers l'imprimante (*printer redirecting*).

Cette deuxième phase pourrait sembler entièrement à la charge des logiciels d'émulation et des résidents, mais elle peut être plus ou moins facilitée par la façon dont se déroule et se conclut la phase 1 de la recherche, et selon que le logiciel de l'hôte est plus ou moins bien 'disposé' à l'égard de la réalisation des phases 2, 3 et 4. En d'autres termes, la préparation de la sortie de la part du système propriétaire, dans ce cas Dobis, influence la capture des données: combien d'entrées d'accès permettant de repérer les données sont-elles disponibles dans la recherche booléenne? Comment les données se présentent-elles sur l'écran? Est-il possible de sauvegarder une série de références comme résultat d'une recherche? Est-il possible de trier ces données? Dispose-t-on d'une commande directe d'exportation des données? etc.

3 Manipulation et conversion, *postprocessing*

Cette phase peut comprendre à son tour deux passages, alternatifs ou complémentaires:

- reformatage permettant de confectionner un produit complet; on l'appelle également remodelage (*repackaging*). Il s'agit du travail effectué, le plus souvent grâce à des macro-instructions du traitement de texte, pour le traitement des données repérées au cours d'une recherche, parfois à partir de plusieurs sources et en plusieurs formats. Il comprend: collation, nettoyage, homogénéisation, élimination des doublons, mise en évidence, classement, commentaires... Lorsque le but est de présenter un document écrit, comme un rapport de recherche destiné à

⁷ "... urged all database publishers to do some rapid policymaking about the downloading of data for storage on local microcomputer", cf. "Cuadra calls for ground-rules on downloading." *Online review* 7, no. 1, (1983): 8.

⁸ Presque tous les projets de rattrapage du passé, pour des fichiers documentaires consistants, prévoient une phase de chargement automatique des données.

⁹ Spore, Stuart. "Downloading from the OPAC: the Innovative Interfaces environment." *Library Hi Tech* 9, no. 2 (1991): 69-79.

- obliger l'utilisateur final, désirant créer des fichiers dérivés, à saisir lui-même toutes les données au clavier. Métaphoriquement, ce serait comme indiquer, au lieu du pont entre deux hauteurs, le gué situé au fond de la vallée.

On pourrait voir dans ces considérations l'expression d'un état d'alerte et d'une urgence, pourtant le téléchargement et la réutilisation des données ne sont pas du tout une nouveauté des années quatre-vingt dix.

En 1982, au cours de la Online Conference d'Atlanta, Carlos Cuadra, Président de la Cuadra Associates, "sollicitait tous les distributeurs de base de données à décider rapidement d'une politique concernant le téléchargement des données et leur stockage dans des micro-ordinateurs locaux"⁷.

Toutefois, en dehors du monde anglo-saxon, le problème a eu une résonance très réduite, et continue à en avoir moins que son pendant: le téléchargement (*uploading*)⁸. En 1991, un article élaboré dans le contexte de la New York University Law Library pouvait encore commencer par ces mots: "Pourquoi télécharger?" ("Why download?")⁹. Si on se demande encore 'pourquoi', c'est parce qu'on peut aussi ne pas être en faveur du téléchargement. Nous reviendrons sur cette question en conclusion.

3 Les phases et les opérations du téléchargement dans le détail

Nous avons déjà dit que la procédure globale comprenant le repérage, la capture et la réutilisation de données d'une ou de plusieurs bases de données centrales vers des bases de données habituellement plus petites est appelée téléchargement. En réalité, les différentes phases sont variées, et le téléchargement n'est qu'une d'entre elles. On peut distinguer au moins les phases suivantes:

- 1 interrogation et repérage
- 2 capture et copie
- 3 manipulation et conversion
- 4 chargement et réutilisation

1 Interrogation et repérage

Démarrage du micro-ordinateur, lancement des logiciels résidents en mémoire TSR (*terminate and stay resident*), du logiciel d'émulation d'un terminal vidéo, VDU (*video terminal unit*). Accès à l'hôte et à l'application. Opération de recherche et de repérage des données.

2 Capture et copie, déchargement des données

Prélèvement des données un écran après l'autre (*screen-by-screen*), capture d'une session d'interrogation (*log file*), ou bien capture d'une série de références formant un fichier. Cela en interceptant les données, soit sur écran, soit grâce à une commande d'enregistrement direct du fichier sur disque, ou après qu'elles ont été dirigées vers l'imprimante (*printer redirecting*).

Cette deuxième phase pourrait sembler entièrement à la charge des logiciels d'émulation et des résidents, mais elle peut être plus ou moins facilitée par la façon dont se déroule et se conclut la phase 1 de la recherche, et selon que le logiciel de l'hôte est plus ou moins bien 'disposé' à l'égard de la réalisation des phases 2, 3 et 4. En d'autres termes, la préparation de la sortie de la part du système propriétaire, dans ce cas Dobis, influence la capture des données: combien d'entrées d'accès permettant de repérer les données sont-elles disponibles dans la recherche booléenne? Comment les données se présentent-elles sur l'écran? Est-il possible de sauvegarder une série de références comme résultat d'une recherche? Est-il possible de trier ces données? Dispose-t-on d'une commande directe d'exportation des données? etc.

3 Manipulation et conversion, *postprocessing*

Cette phase peut comprendre à son tour deux passages, alternatifs ou complémentaires:

- reformatage permettant de confectionner un produit complet; on l'appelle également remodelage (*repackaging*). Il s'agit du travail effectué, le plus souvent grâce à des macro-instructions du traitement de texte, pour le traitement des données repérées au cours d'une recherche, parfois à partir de plusieurs sources et en plusieurs formats. Il comprend: collation, nettoyage, homogénéisation, élimination des doublons, mise en évidence, classement, commentaires... Lorsque le but est de présenter un document écrit, comme un rapport de recherche destiné à

⁷ "... urged all database publishers to do some rapid policymaking about the downloading of data for storage on local microcomputer", cf. "Cuadra calls for ground-rules on downloading." *Online review* 7, no. 1, (1983): 8.

⁸ Presque tous les projets de rattrapage du passé, pour des fichiers documentaires consistants, prévoient une phase de chargement automatique des données.

⁹ Spore, Stuart. "Downloading from the OPAC: the Innovative Interfaces environment." *Library Hi Tech* 9, no. 2 (1991): 69-79.

l'utilisateur, le téléchargement peut s'arrêter ici. Mais quand le but - comme dans notre cas - est la réutilisation des données dans des fichiers cibles, il ne peut s'agir que d'une phase complémentaire qui sert au limage préliminaire des aspérités dans les données provenant du système source

● reformatage comme conversion des données pour l'importation et la réutilisation dans un système local (SGBD, feuille électronique, ISRS, BFS). Il implique une phase séparée de préparation et de test hors ligne du profil de conversion, suivie de l'exécution effective. Son fonctionnement est influencé par le contenu et le format des références capturées dans la phase 2.

La conversion peut:

- a) être incorporée dans le système qui fournit les données [ex.: *INNOPAC*, *LBSPC/Libis-Net* (Leuven)];
- b) concerner un logiciel intermédiaire [ex.: *Fangorn*, *Biblio-link*, *Data Magician*, *Refwriter*, *EndLink* de *EndNote*, *Capture de Reference Manager*];
- c) faire partie intégrale du système cible doté de fonctions d'importation et de filtrage [ex.: *Papyrus*, *Library Master*, *Idealist*, *CDS/ISIS* (Reformatting FST¹⁰)].

4 Chargement et réutilisation

L'exploitation finale dépend de l'environnement cible où sont exploitées les données prélevées.

4 Téléchargement des données bibliographiques en milieu Dobis: expériences documentées

Qu'a-t-on déjà réalisé en milieu Dobis?

- 1 DBD et MICROBIS (Université de Modène) (Panza 1989)
- 2 Biblio-link PBS/Pro-Cite (Emory University, Atlanta) (Deemer, 1991)
- 3 LBSPC/Libis-Net (Leuven) (Vanautgaerden, 1991)
- 4 Textsearch, Datamat Rome (ENEA-Italie) (Benadusi et al. 1992)¹¹.

Que nous offrent les réalisations déjà effectuées en milieu Dobis?

- 1 Dobis a été parfois profondément modifié, au point que la réalisation n'est pas proposable à d'autres utilisateurs¹²
- 2 d'autres réalisations sont des approches externes, sans modification de Dobis, mais elles ont été effectuées à partir de programmes réalisés ad hoc, qui ne circulent pas comme produits, et qui sont parfois des prototypes. Il s'agit d'applications réalisées par des informaticiens, sujettes à manutention à chaque changement littéral, et qui ne sont pas modifiables par les bibliothécaires
- 3 l'approche Pro-Cite ne modifie pas Dobis, n'exige pas de programmation et est entièrement entre les mains des bibliothécaires. Mais le résultat a plusieurs défauts - nous les verrons plus avant -; il s'agit d'un produit limité à une seule expérience, et la personnalisation en est plutôt réduite.

5 Notre intervention

5.1 Objectifs et orientations

Nous avons donc considéré que les réalisations existantes autorisaient des tentatives supplémentaires, et surtout différentes. Voici les limites à l'intérieur desquelles nous avons inscrit notre intervention, fixé nos buts et évalué les résultats:

¹⁰ FST Field select table est une table utilisée par Micro CDS/ISIS de différentes façons, par exemple pour filtrer et manipuler les données importées et exportées en format ISO 2709: Fangorn convertit en ISO 2709, de sorte que les données sont ensuite modifiables par la FST qui dispose d'instructions particulières, par ex. IF, avec contrôle croisé des zones, traitement articulé des sous-zones, etc. Pour le format ISO 2709 cf. la norme du même nom "Disposition des données sur bande magnétique pour l'échange d'informations bibliographiques" déc. 1987.

¹¹ Cf.: Panza, Adriana. "Automated Dobis sessions: a first attempt to use a PC as an inter-Dobis communication tool." *8th Annual Meeting of The Dobis Users Group*, 5-8 September 1989, Liverpool, Great Britain. Deemer, Selden. "PC bibliographies from Dobis records." *Bulletin Dobis-Libis = Dobis-Libis newsletter* 9, no. 3 (December 1991): 4-6. Vanautgaerden, Jos. "Dobis downloading." *10th Annual Meeting of The Dobis Users Group*, 10-13 September 1991, Tokyo, 85-95. Benadusi, Marco, Angelo Colombini, Alessandra Ensolì, Giorgio Grimaldi, Gianluca Rispoli. *Elaborazione e diffusione della letteratura tecnica in ENEA*, 31 p., [29] c., [Roma], 1992.

¹² Communication personnelle entre F. Dell'Orso et Jos Vanautgaerden, Leuven 10 décembre 1991: «...to implement this requires a lot of prerequisites. The LEUVEN downloading process not only involves lots of changes to existing programs and a number of new ones, but it also needs profound customizing to meet your specific (installation) needs».

- 1 examiner ce qu'il est possible de faire avec le Dobis actuel, jusqu'à l'année 1995 comprise - car nous ne pensons pas que nous aurons son "successeur" avant la Noël de l'année en question - en vérifiant la faisabilité de deux façons:
 - I *sans aucune modification* de Dobis et, sans écrire de nouveaux programmes
 - II *en modifiant* Dobis, mais uniquement là où il a été déclaré paramétrisable, c'est-à-dire dans la Description complète de l'OPAC LIBI, recherche simplifiée pour le public, (programme DBSSE11) et dans l'impression d'une liste de documents sauvegardés (programme DBSSE09)
- 2 chercher une façon de procéder qui soit placée sous le contrôle de l'utilisateur final (bibliothécaire, lecteur, etc., sachant utiliser un micro-ordinateur), et qui lui permette:
 - a) d'agir tout seul, sans recourir à des programmes ou à des interventions informatiques différents de ceux qui sont déjà disponibles comme produits finis;
 - b) de choisir de façon autonome temps et lieux pour opérer;
 - c) de maintenir l'application dans le temps en l'adaptant en fonction des différents types de données et des changements qui peuvent se produire à la source, indépendamment de son contrôle, sans devoir modifier de programmes
- 3 utiliser des logiciels accessibles à tous, disponibles sur le marché, commerciaux ou non, distribués dans le monde entier, et à bas prix.
 Nous nous sommes orientés vers des logiciels qui fonctionnent en système d'exploitation MS-DOS:
 - émulateurs du type 3270
 - *printer redirecting utilities*: petits logiciels utilitaires permettant de dévier les données, en sortie sur l'imprimante
 - logiciels utilitaires résidants en mémoire: TSR
 - traitements de texte
 - logiciels de conversion
 - gestionnaires de fichiers documentaires tels que l'utilisateur ait la possibilité de personnaliser et de définir la structure de la base de données sans recourir à des langages de programmation, et qui soient pourvus de possibilités d'importation
- 4 vérifier la circulation des données entre les différents environnements cibles: ou, en d'autres termes, la capacité d'import/export entre les logiciels documentaires dans lesquels les données sont chargées et entre ceux-ci et les traitements de texte, les logiciels statistiques, les SGBD relationnels
- 5 nous concentrer surtout sur les phases 2, téléchargement, et 3, manipulation, en contexte Dobis standard
- 6 au vu des résultats, formuler: des observations sur les difficultés que l'on rencontre actuellement, des propositions pour le nouveau Dobis, des suggestions pour son successeur
- 7 envisager de toute façon une 'solution informatique', en mesure de dépasser les limites de cette orientation, et à rechercher toujours dans une approche extérieure à Dobis, sans modification des programmes¹³.

5.2 Objectifs à négliger

Vice versa, voici de quoi nous avons décidé de ne pas nous occuper, à part, naturellement, ce qui pourrait s'opposer aux objectifs précédemment formulés.

- 1 analyse de la recherche en Dobis comme fonction et formulation d'une proposition homogène d'amélioration ou de remaniement
- 2 données non bibliographiques (administratives, statistiques, etc.); données bibliographiques différentes des descriptions catalographiques complètes, par exemple: entrées d'un fichier d'autorité (ex.: vedettes matière); données bibliographiques se rapportant à une référence catalographique, mais non incluses dans la Description complète ou non imprimées dans la liste des documents sauvegardés (relations, renvois, résumés, données bibliographiques locales...)
- 3 systèmes receveurs du genre des feuilles électroniques (*spreadsheets*)
- 4 copyright
- 5 analyse et appréciation des caractéristiques des différents systèmes receveurs, traités ici plus comme une famille que comme des logiciels individuels à choisir
- 6 manipulation limitée au seul remodelage du texte; reformatage direct uniquement par traitement de texte¹⁴.

¹³ Cf. § 7 et la note 43 p. 18.

¹⁴ On pourra obtenir ce dernier à partir des systèmes receveurs une fois les données transférées, le produire directement par une opération de remodelage donnerait des données de sortie déjà définies, orientées, rigides.

5.3 Logiciels utilisés

5.3.1 Logiciels d'émulation, de capture des données, de traitement de texte

- émulateurs 3270: IBM Personal Communication 2.0, IBM Emulation Program 3.05
- *printer redirecting utilities* et autres logiciels utilitaires résidants en mémoire *TSR* (Windows Microsoft, Sidekick Borland, PRN2FILE.COM, LPT2FILE.COM¹⁵)
- Wordperfect 5.1.

5.3.2 Logiciels de reformatage et de conversion

Nous avons testé des logiciels¹⁶ offrant des fonctions de conversion et d'importation de façon différente:

- a) incorporées dans le système cible, orientées vers un système et un format définis, bien que personnalisables: on y trouve des logiciels orientés *un-pour-un* (Pro-Cite & Biblio-link avec Dobis), *plus d'un-pour-un*, où le 'plus d'un' peut être à définir (Library Master)
- b) situées dans un logiciel ad hoc, séparé du système cible; orientées pour ne produire que le format du système receveur, mais disposées à recevoir n'importe quel format à intitulés ('tagged': à étiquettes, à labels) et le format avec séparateur et délimité comme entrée: *de-n-à-un* (Fangorn pour le format ISO 2709 de CDS/ISIS)
- c) logiciels entièrement consacrés à la conversion du format des données bibliographiques: ce sont des logiciels généraux, parfois universels, qui acceptent de nombreux formats en saisie et qui peuvent produire de nombreux formats en sortie *de-plus-d'un-à-plus-d'un* (Data Magician) *de-n-à-n* (Refwriter)¹⁷.
- d) tous autorisent une importation et permettent une exportation en format avec séparateur et délimité (ex.: *comma delimited*), tous peuvent exporter sur disque comme simple texte ASCII, sans caractères de contrôle.

5.3.3 Logiciels de gestion de fichiers

Nous avons testé comme systèmes receveurs deux logiciels qui appartiennent à la famille des *BFS Bibliography formatting software* (logiciels de formatage des bibliographies), et un de ceux du groupe des *Information storage and retrieval systems* (logiciels de mémorisation et de dépistage des informations)¹⁸. Nous avons en outre vérifié le transfert des données en des SGBD comme dbIII+ (Ashton Tate), dans des logiciels statistiques comme EPI INFO et dans des traitements de texte différents.

Les programmes de gestion documentaire pris en considération se prêtent à la gestion de bibliographies et de catalogues personnels, et aussi d'archives de bibliothèques contenant des dizaines de milliers de références. Ils sont tous dotés des fonctions de saisie, correction, maintenance, sortie, recherche, sauvegarde, exportation.

Les BFS ont la particularité d'offrir des batteries d'outils déjà prêts pour le travail bibliographique:

- a) formats prédéfinis pour la mémorisation du matériel: différents bordereaux permettant d'entrer les données concernant les livres, les articles, les rapports, les audio-visuels, les manuscrits, la musique, etc.
- b) formats prédéfinis pour la présentation des données en sortie (écran, disque, papier), selon les exigences de style requises par les différents standards utilisés par les revues et les maisons d'édition (styles comme: ANSI, Science, Nature, MLA, APA, Turabian, Chicago, etc.).

Tous les logiciels satisfont aux conditions indiquées au § 5.1, comma 3, p. 5.

¹⁵ PRN2FILE est un logiciel distribué par *PC Magazine* en 1987; LPT2FILE est réalisé par J. Prosise et est inclus dans son volume *DOS 5, Techniques and utilities*, 1991. Les deux programmes sont à peu près équivalents: nous ne citerons que le premier.

¹⁶ Pour les références aux organismes propriétaires cf. § 8 p. 19.

¹⁷ Data Magician traite en entrée et en sortie: ASCII délimité, avec intitulé (*tagged*), dbase, INMAGIC, MARC Communication format, STAR, Micro CDS/ISIS. Refwriter n'a été utilisé que dans la version démonstration, et n'a pas été réellement testé.

¹⁸ Dans la famille des *BFS*, on compte plus de 50 produits. Parmi eux: Pro-Cite, Idealist, Library Master, Reference Manager, EndNote, Papyrus. Parmi les *ISRS*, on compte également: Texto Logotel, Microquestel, Inmagic, Micro CDS/ISIS. Ce sont tous des systèmes pour la création de bases de données textuelles sur micro-ordinateur en contexte DOS. Beaucoup existent aussi en version Macintosh, et avec compatibilité OS/2, Windows.

5.4 Déroulement de l'expérience

Nous avons utilisé des données issues de catalogues Dobis en ligne des installations suivantes: Emory University, Oxford University-OLIS, ENEA Rome, et surtout Université de Modène et de Pérouse. Les données ont été capturées sur écran. Comme nous l'avons déjà dit - § 5.1 p. 4 - nous avons expérimenté deux approches: la première sans modification de Dobis, l'autre avec modification. Les données ont été capturées sur écran, et dans le cas de Pérouse et de Modène, en interceptant la liste de documents sauvés acheminée vers l'imprimante: ce dernier mode a fonctionné en local et aussi en télétraitement.

5.4.1 Approche sans modification de Dobis

La «Description complète» de Dobis - gérée par le programme DBSCA03 - présente l'avantage d'être vraisemblablement identique dans toutes les installations Dobis, et donc d'être un standard, tandis que, dans la même mesure, la Description complète de l'OPAC LIBI (recherche simplifiée) est probablement différente selon les installations, et il nous a fallu à nous aussi rapidement la modifier.

Dans la Description complète de Dobis, nous avons relevé ces caractéristiques:

- 1 absence de données: zones fixes, zones d'ouverture, données bibliographiques locales et données sur l'exemplaire (à part relations, renvois et résumés)
- 2 indifférenciation¹⁹: toutes les entrées de fichiers point d'accès sont sur le même plan, en ordre de saisie, au-delà de la forme des vedettes (type de nom ou de titre) et du type d'entrée (principale, secondaire); les sous-zones (ex. noms et éditeurs) ne sont marquées que par la ponctuation entrée. Au contraire, les noms sont accompagnés de l'indication de la fonction auteur. Les notes ont une étiquette qui les identifie comme telles, mais en position variable (pas en deuxième colonne, comme les autres zones, cf. la note 39 p. 16)
- 3 quatre-vingtième colonne: les chaînes sont interrompues à la 79ème colonne, la 80ème est occupée par un retour de chariot (ASCII <13> <10>) qui se traduit par un blanc (ASCII <32>) lorsque la chaîne continue à la ligne suivante
- 4 données dépourvues d'étiquettes d'identification: dates de publication, numéro de volume dans la collection, (pour les entrées de fichier au sein du même groupe, cf. supra, 2)
- 5 'f bas' ('forward'): les données qui n'entrent pas sur un écran sont visualisées sur l'écran suivant qui a les mêmes en-têtes; on a aussi redondance de données
- 6 le caractère qui sépare les occurrences d'un champ répétitif, (i.e. multi-articles) c'est-à-dire ' / ' ou ' / ' peut faire partie du contenu de la zone, ex.: Bulletin Dobis/Libis / Dobis/Libis Newsletter
- 7 les entrées sont visualisées uniquement sous la forme d'affichage: pour visualiser les lettres accentuées il faut utiliser celles qui sont disponibles sur les touches du clavier (ce qui est notoirement limité et entraîne d'autres problèmes)
- 8 les éléments descriptifs ne sont pas présentés dans une succession conforme aux formats ISBD et AACR2.

Ces caractéristiques ont été considérées comme une limite: comment ont-elles été affrontées?

- 1 absence de données: rien à faire
- 2 indifférenciation: rien à faire pour les APFE, les entrées et même pas, de façon certaine, pour les sous-champs²⁰; les notes ont une étiquette individuelle: on peut donc les extraire à partir d'une macro-instruction de traitement de texte²¹

¹⁹ Dans de nombreux systèmes, parmi lesquels Dobis, basés sur un riche encodage des données - dans notre cas du type MARC -, on relève une disproportion macroscopique entre ce qui est mémorisé à partir de la saisie, et les données de sortie sur papier ou sur écran. Puister directement dans le format interne des données permet de les récupérer intégralement (cf. § 7 p. 18). De plus, si généralement ce qui a été distingué en entrée tend à se confondre en sortie, les macro-instructions des éditeurs et des traitements de texte, utiles, sinon nécessaires, pour nettoyer et classer les données de sortie, et les programmes d'importation, auront alors bien d'autres difficultés qui ne se résoudront pas facilement sans instructions conditionnelles: 'IF...THEN...ELSE'.

²⁰ En effet, les sous-zones sont montrées en sortie uniquement avec la ponctuation entrée par le catalogueur. Si celle-ci coïncidait de façon univoque avec les sous-zones, des logiciels comme Fangorn, Data Magician et, dans la seule zone de l'éditeur, Pro-Cite, pourraient l'intercepter et la traiter, mais ce n'est pas le cas. Voir ces exemples: Bismarck, Otto, Fürst von (AACR2r 22.12) Smith, John, 1924- (AACR2r 22.17) Madrid : Ayuntamiento, Secretaria General (Reglas 2.4.4D). Ainsi la ponctuation utilisée comme marque d'une sous-zone un nombre indéterminé de fois, mais pas la première fois lorsqu'elle coïncide avec le début du champ, met ces logiciels en difficulté: New York ; London ; Sidney : Oxford University Press (AACR2r, 1.4.D1) Paris : Gauthier-Villars ; Chicago : University of Chicago Press (AACR2r 88, 1.4.D5) London : MacMillan : Educational Service [distributor] (AACR2r, 1.4.E1).

²¹ Les étiquettes de la fonction auteur peuvent devenir des leviers de commande: aussi bien Fangorn, et surtout la Reformatting FST d'Isis, que Data Magician peuvent les intercepter et les utiliser pour les orienter vers des zones ou des sous-zones particulières. Tout aboutit normalement si les indications de la fonction auteur ont été utilisées en entrée de façon instrumentale, pour standardiser les en-têtes, et non littéralement pour décrire les fonctions des noms selon le MARC. Il est paradoxal que l'on puisse utiliser la donnée la moins importante, mais c'est une démonstration

- 3 blanc dans la 80ème colonne: peut être éliminé par une macro-instruction de traitement de texte
- 4 manque d'étiquette d'identification: rien à faire, de façon certaine et non approximative, pour les dates²², et pour le numéro de volume dans la collection
- 5 'f bas' ('f forward'): une macro-instruction de traitement de texte permet d'intercepter les références qui s'alignent sur plus d'un écran: on élimine l'en-tête de l'écran suivant identique à celui de début de référence, on interpose une zone qui servira de signal avisant de la présence de données redondantes, à éliminer manuellement de toute façon, dans le système cible; rien à faire pour le reste
- 6 séparateur d'occurrences: 'blanc/blanc', ou 'blanc/', peut se trouver contenu dans les zones; on recourt à une commande de "cherche et remplace" non globale, mais à confirmer à chaque article
- 7 forme diacritique: rien à faire, on capture ce qui s'affiche sur l'écran
- 8 ordre des éléments: rien à faire.

De plus, sans aucune modification de Dobis, la capture des données ne peut se faire que par copie d'écran, soit d'un seul, soit d'une séquence, mais non par l'interception de la liste de documents sauvés, car le format de visualisation utilisé dans celle-ci est dans tous les cas moins fourni que la Description complète standard.

5.4.2 Approche avec modification de Dobis

Nous avons décidé de modifier non la Description complète Dobis, mais celle de l'OPAC LIBI sur écran - déjà conçue de façon tabellaire en Dobis, pour permettre justement de l'adapter aux exigences locales, ainsi que le programme d'impression d'une liste de documents sauvegardés: aucune intervention sur le format interne des données, et cela en raison des objectifs déflucidés au § 5.1 p. 4.

La Description complète a été conçue sur le modèle des formats propres à de nombreuses bases de données sur CD-ROM et sur services en ligne comme un format étiqueté, à intitulés (*variable tagged format*), qui est, tout compte fait, déjà le format d'origine Dobis.

de la façon dont fonctionnent les reformatages: un élément distinctif est nécessaire, après quoi il devient un point d'appui pour une manipulation efficace. Par contre les indications de la fonction auteur peuvent être éliminées par une macro-instruction de traitement de texte.

²² Dans la Description complète standard de Dobis, les chaînes et, parmi elles, celles qui sont formées de nombres, comme les dates, peuvent être interrompues par un <CR> après la 79ème colonne. Éliminer tous les <CR> compromet la lisibilité des étiquettes et des autres zones. Pour les dates, cela veut dire que le groupe considéré comme distinctif 'blanc-1xxx-<CR>-prolongement sur la ligne suivante de 5 blancs' [à répéter de façon différente pour chaque type de date: multiple, incertaine, copyright, réimpression etc.] peut être interrompu à chaque stade de la séquence jusqu'à <CR>. Il faudrait une macro-instruction pour chaque combinaison possible. Il faut avant tout éliminer les séquences standard des 16 blancs, car ce groupe peut se présenter aussi dans la zone du titre et des notes, en provoquant de la sorte une hécatombe de données. Si l'on se contente d'un résultat approximatif et imprécis, certaines références pourraient rester sans date. Seul Data Magician sait relever la situation en émettant un signal. Mais à ce moment-là, comment retrouver la date? Effectuer le téléchargement de toute la partie descriptive 'supérieure' de la Description complète, pour l'inclure, semble un peu encombrant.