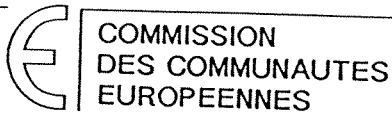


JM



COMMISSION  
DES COMMUNAUTÉS  
EUROPÉENNES

Version octobre 1991

DIRECTION INFORMATIQUE  
BASES DOCUMENTAIRES  
Celex

# MODERNISATION

## CELEX

Sous-projet 2  
Système d'Alimentation CElex - ARCHIVE  
(production de la partie analytique) et  
Bases de diffusion (BDD)

## ALICE - ARCHIVE/BDD

ETUDE DE FAISABILITE

(Projet)

## TABLE DES MATIERES

- 0. Résumé du contenu
- I. Définition du problème et situation actuelle
  - I.1 Introduction
  - I.2 La Modernisation de CELEX
  - I.3 Situation actuelle
- II. Définition des objectifs
  - II.1 Amélioration du noyau
  - II.2 Amélioration de l'accès du système
- III. Spécifications utilisateur
- IV. Structure du système d'information : solution proposée et solutions alternatives
  - IV.1 Contexte général du système
  - IV.2 Architecture de la solution proposée : description fonctionnelle
  - IV.3 Architecture informatique (configuration hw/sw)
  - IV.4 Multilinguisme
  - IV.5 Relation entre les sous-systèmes ALICE-T et ALICE-A
  - IV.6 Solutions alternatives
  - IV.7 Evolution technique
  - IV.8 Aspects sécurité
- V. Développement et mise en place
  - V.1 Stratégie de développement
  - V.2 Stratégie de mise en place
  - V.3 Plan de migration de l'ancien système
  - V.4 Estimation des ressources
- VI. Aspects Organisationnels
- VII. Analyse coûts/bénéfices
- VIII. Responsabilités et services participants
- IX. Facteurs de risque
- X. Plan du projet

Annexes :

## 0. RESUMÉ DU CONTENU

La modernisation de CELEX, en suspens depuis 1985, a été lancée fin 1988 par l'adoption de la préanalyse. Le projet global a été décomposé en deux sous-projets:

sous-projet 1 : Production des données textuelles (Texterface) dont la construction a commencé en Juillet 1989 suite à l'adoption de l'étude de faisabilité;

sous-projet 2 : Production des données analytiques, bases de diffusion et "pilote" (Archive/BDD), qui fait l'objet de la présente étude de faisabilité.

L'objectif du projet est principalement:

**« à rendre l'accès au droit communautaire  
le plus convivial possible ».**

Cet objectif qui est imposé à CELEX par des résolutions du Parlement européen, par le Groupe Informatique juridique du Conseil, par le COREPER et surtout par les quelques 2.800 utilisateurs du système, doit être atteint à tout prix afin de garantir la meilleure diffusion du droit communautaire dans le cadre de l'objectif 1993 et au-delà.

Plus particulièrement le présent sous-projet 2 vise d'abord à l'amélioration du noyau de l'application (consolidation et rationalisation de la structure logique des documents et de la base, extension de la couverture, accélération des mises à jour et amélioration de l'actualité, facilité de gestion etc.). Le sous-projet 2 vise aussi à amélioration de l'accès au système (introduction du CCL, assistance en ligne, possibilité de téléchargement, par l'introduction de caractères riches, document delivery, prévision d'un didacticiel etc.) mais surtout l'introduction d'une version coordonnée des textes modifiés.

La solution proposée, constitue un élargissement de la solution adoptée dans le cadre du 1er sous-projet (TEXTERFACE); à son noyau se trouve une base de données relationnelle (ORACLE) sur une machine locale UNIX. L'échange des données entre le système local et le système central ainsi qu'avec l'OPOCE et les autres services et institutions se fait à l'aide des formats standardisés du type SGML, ce qui rend l'application indépendante des technologies (h/w et s/w) utilisées. En plus l'utilisation des éléments basés sur UNIX et le modèle relationnel rendent l'application la plus ouverte et la moins captives que possibles.

Pour réaliser ce projet il est indispensable de mettre à la disposition du project management les ressources suivantes:

Personnel (15,7 H/A):

7,5 H/A en sous traitance externe (750.000 ECU) sur deux ans

3,8 H/A en sous-traitance Traduction sur 1 an

4,4 H/A ressources internes IX/I (dont 1 B à engager en permanence et le reste sur 2 ans)

Espace disque:

AUGMENTATION NETTE : 18 GB

PENDANT PROJET PREVISION DE 6,5 GB SUPPLEMENTAIRES

ENSUITE PREVISION DE 1 GB PAR AN

Les coûts en personnel pour le projet représentent moins que les 2/3 du coût annuel de CELEX et seront largement récompensés par les bénéfices plus que remarquables:

- accélération des mises à jour de 86 à 93%;
- diminution de 41% du coût de la maintenance;
- diminution de 50% de la formation (qui peut conduire à des économies de 30 à 65% du coût de formation pour les utilisateurs externes);
- économie de 20 à 70% des coûts d'utilisation du système

Dans le cas où le projet n'est pas approuvé, la Commission se verra reprocher de ne pas vouloir soutenir un projet dont la bonne fin est jugée indispensable par les utilisateurs (dont plus de 1.240 clients payants) et les autres institutions, en particulier le Parlement et le Conseil.

## I. DEFINITION DU PROBLEME ET SITUATION ACTUELLE

### I.1. Introduction

CELEX, le système interinstitutionnel de documentation automatisée pour le droit communautaire, a été créé sous la responsabilité du Service juridique de la Commission pour répondre au besoin d'un système d'information performant en la matière. La base de données a été déclarée opérationnelle en janvier 1971.

Dans sa résolution du 26 novembre 1974 relative à l'automatisation de la documentation juridique (JOCE C 20 du 28 janvier 1975), le Conseil a décidé de mettre le système à la disposition des Etats membres et de créer un comité d'experts composé de représentants des Etats membres et des Institutions communautaires. Ce comité, dit "Groupe Informatique juridique du Conseil", a reçu, entre autres, le mandat de suivre les développements du système CELEX.

La Commission assure la gestion du système dont la responsabilité a été confiée à l'ensemble des Institutions. La coordination des travaux s'effectue dans le cadre du Comité interinstitutionnel CELEX.

Le système couvre l'ensemble du droit communautaire, réparti en 10 "secteurs" précisément définis (deux de ces secteurs sont encore à créer). En août 1990, CELEX contient plus de 130.000 documents par version linguistique.

La base de données permet l'accès à une information fiable et précise sur la législation communautaire, la jurisprudence, les travaux préparatoires et les questions parlementaires en utilisant les fonctionnalités d'un progiciel de recherche documentaire de type "texte intégral".

Ouvert successivement aux Institutions puis au public, CELEX est maintenant exploité par plus de 1.200 utilisateurs externes et 1645 utilisateurs internes. Le moule dans lequel le système se présente actuellement est le résultat d'une longue évolution.

Suite à plusieurs conversions, l'ensemble de l'édifice est devenu si complexe - et donc si fragile - que le coût de la maintenance évolutive ou même corrective atteint un seuil prohibitif alors que, simultanément, la gestion quotidienne de l'application devient de plus en plus lourde et difficile.

Le système actuel est arrivé en quelque sorte à ses limites et ne peut pas suivre les rythmes accélérés qui sont nécessaires pour faire face aux défis de l'Europe de l'an 2000.

Il est donc devenu impératif et urgent d'adapter CELEX aux besoins de ses utilisateurs comme de ses administrateurs, ainsi qu'aux standards et concepts des nouvelles technologies de l'information. Le système actuel doit faire l'objet d'une refonte complète permettant la mise en place d'un outil d'information performant pour les dix années à venir.

## 1.2. La modernisation de CELEX

Le projet de modernisation CELEX a été lancé par l'adoption de la "Préanalyse pour le développement et la modernisation du système CELEX" (réunion du 30.11.1988 v. ref. 1). Le projet global de développement a été décomposé en deux sous-projets :

sous-projet 1 : Production des données textuelles c.à.d. construction de la base de production des textes et du "texterface" (interface d'alimentation des textes); l'étude de faisabilité (ref. 2) a été adoptée en réunion le 7.7.89 (doc. IX/1/4 - (89) D/214 du 14.7.89);

sous-projet 2 : Production des données analytiques, bases de diffusion et administration, c.à.d. construction de la base de production des données analytiques (BDP-Archive), des bases de diffusion, de l'interface entre production et diffusion ainsi que du module administration.

La présente étude de faisabilité concerne le sous-projet 2.

## 1.3. Situation actuelle

### a. Les bases

Tel qu'il se présente à l'utilisateur, les principales caractéristiques de CELEX peuvent être résumées comme suit :

- 131.451 unités documentaires (au 9 août 1990), réparties sur 8 des 10 secteurs prévus. La répartition par secteur étant:

* secteur 1 (Traités)	: 2.400
* secteur 2 (Accords)	: 1.734
* secteur 3 (Droit dérivé)	: 41.754
* secteur 4 (Droit complémentaire)	: 487
* secteur 5 (Travaux préparatoires)	: 14.493
* secteur 6 (Jurisprudence)	: 5.632
* secteur 7 (Mes. nationales d'exécution):	17.688
* secteur 9 (Questions parlementaires)	: 47.262

(La couverture prévue en textes intégraux est présentée dans l'Annexe I de l'étude de faisabilité TEXTERFACE).

- 31 lexiques, répartis sur 80 champs utiles (voir annexe 1A de la préanalyse)

- mises à jour hebdomadaires (voir statistiques en annexe 1B de la préanalyse); les délais de mise à jour sont les suivants:

- \* document publié dans le JO ou document COM:
  - . partie analytique 2-3 semaines après publication;
  - . texte 4-5 semaines après publication;

- \* arrêt de la Cour de Justice:
  - . partie analytique 4 semaines après prononcé
  - . texte 8 semaines à 3 ans après prononcé selon langue.

- disponibilité en français, anglais, allemand, néerlandais, italien, et danois; versions grecque et espagnole en pilote; version portugaise en préparation. Les versions linguistiques autres que française font l'objet d'une "couverture différenciée", qui donne lieu à un filtrage de certaines données avant chargement (voir annexe 1E de la préanalyse).

Les versions disponibles peuvent être interrogées en mode TTY via une grande variété de terminaux (voir annexe 1C de la préanalyse). Depuis Mars 1990 les données textuelles nouvellement introduites dans les bases latines contiennent des caractères spéciaux (caractères "riches") selon la norme ISO 8859/1. Les utilisateurs ne disposant pas d'un terminal supportant cette norme peuvent visualiser les caractères "riches" proprement appauvris en lettres majuscules (caractères "pauvres"). Il n'est pas possible, à l'heure actuelle, de reproduire à l'écran certains tableaux complexes ni les schémas inclus dans les textes.

#### b. La production

Un diagramme des différentes composantes du système, tel qu'il se présente à l'équipe responsable de l'administration de l'application, figure en annexe 1D de la préanalyse. L'introduction des données s'effectue par les chaînes de traitement et d'alimentation qui sont:

- chaîne BIBLIO (données analytiques);
- chaînes TEXTES (IC1, IC2, PER, MOS, INT, EUR, LUX, PHI, SZ, RPE);
- chaîne ANNULATIONS
- passerelle "MESURES NATIONALES D'EXECUTION" (ASMODEE vers CELEX);
- chaîne SYNONYMIE;
- chaîne INFOCELEX.

Le noyau du système de production est un fichier maître connu dans le jargon de CELEX comme l'ARCHIVE. Il s'agit d'un fichier séquentiellement indexé à partir duquel sont produites les données analytiques dans les différentes versions linguistiques de CELEX.

Une description détaillée de chacune des chaînes figure en annexe 1E de la préanalyse.

On peut à titre indicatif, rappeler que l'environnement informatique de CELEX comprend actuellement:

- 122 programmes
- 83 procédures
- 74 JCL's
- 116 fichiers (dont 68 destinés à la diffusion par serveurs)

Un autre point faible de CELEX est son intégration partielle dans l'environnement bureautique qui débute dans les services de la Commission. Actuellement pour faire un téléchargement ("downloading") des textes repris dans CELEX sous forme revisable

en local, l'utilisateur est confronté à une procédure peu conviviale et très peu performante (v. Annexe II de l'étude de faisabilité TEXTERFACE). Les utilisateurs externes font face aux mêmes problèmes.

Pour le moment, il n'existe pas de procédure fiable de "uploading" des fichiers d'une machine locale vers CELEX. L'installation du MFTS sur la machine BULL GCOS 8 sera un premier pas pour pallier cette situation.

Enfin, la gestion de l'introduction des textes est très pénible tandis que leur correction dans les bases est à l'heure actuelle quasiment impossible.

### c. L'exploitation

En termes d'utilisation, les différentes versions linguistiques de CELEX totalisent le taux d'interrogation le plus élevé des bases documentaires de la Commission. Une ventilation du temps d'utilisation de CELEX par les utilisateurs internes et externes figure en Annexe 1.

Outre l'interrogation en ligne, CELEX est utilisé pour :

- la production de "profiles utilisateurs" (une vingtaine par semaine); les services de la Commission se servent de ces profils pour mieux contrôler la législation pour laquelle ils ont la responsabilité;
- l'extraction de données pour les sites serveurs externes (près d'une vingtaine en février 1990);
- l'édition du "Répertoire de la Législation communautaire en vigueur";
- la production du "système d'alerte qui permet aux services de la Commission d'éviter l'apparition de vides juridiques";
- l'alimentation de la base ASMDEE (contrôle de l'application des directives par les Etats membres).



d. Les ressources utilisées

Pour la production, l'exploitation, la maintenance et la modernisation du système CELEX, la Commission et les autres Institutions ont consommé en 1989(1)

Commission :	22 H/A	953.723 ECU	8 GBytes
Cour de Justice :	3,5 H/A	40.000 ECU	
Parlement :	2,5 H/A	15.000 ECU	
Conseil :	1 H/A		
CES :	0,5 H/A		
<b><u>TOTAL :</u></b>	<b><u>29,5 H/A</u></b>	<b><u>1.008.723 ECU</u></b>	<b><u>8 GBytes</u></b>

Il faut signaler ici que la gestion nécessite quelques 5,5 GB avec lesquelles il est possible d'effectuer les m.à.j des 2/3 bases en parallèle; cet espace disque est utilisé en commun avec d'autres bases sous le progiciel MISTRAL.

La Commission gère le système global pour le compte des autres Institutions. En 1991 le détail des ressources se lit :

	<u>Ressources internes</u>	<u>Ressources externes</u>
- Management	13 H/A	
. analyse documentaire		180.000 ECU
. formation utilisateurs	1 H/A	
. encodage textes		712.723 ECU
- Diffusion	5 H/A	
- Maintenance, développement, modernisation	1 H/A	70.000 ECU
- Gestion centrale	2 H/A (sous contrat externe)	
<b><u>Total :</u></b>	<b><u>22 H/A</u></b>	<b><u>962.723 ECU</u></b>

En résumé, la Commission gère avec un effectif de 22 fonctionnaires, un système qui occupe 13,5 GBytes. Ces 22 H/A forment et assurent également le suivi de plus de 3.000 utilisateurs et une vingtaine de serveurs. Il faut aussi signaler ici que l'équipe DBM CELEX tourne actuellement avec un déficit de 5 H/A (4 LAs et 1 B) en personnel.

Il résulte de ces chiffres que la Commission emploie 1,62 H/A par GByte et 1 H/A pour le suivi de 130 utilisateurs.

Il faut aussi mentionner ici qu'un cours de formation CELEX dure actuellement 2 jours.

(1) Rapport de la Commission au Groupe informatique juridique du Conseil.

## II. DEFINITION DES OBJECTIFS

Le projet de modernisation du système CELEX s'attache avant tout:

« à rendre l'accès au droit communautaire le plus convivial possible ».

Pour arriver à ce but trois classes d'objectifs doivent être atteints:

- a. amélioration du noyau de l'application (structure logique, couverture, actualité, facilité de gestion etc.)
- b. amélioration de l'accès au système (interfaces conviviaux, CCL, assistance en ligne, téléchargement, document delivery etc.)
- c. mettre en place une organisation assurant l'amélioration de méthodes de saisie/encodage/transmission des textes introduits dans les bases.

Plus particulièrement les objectifs à atteindre sont les suivants:

### II.1 Amélioration du noyau

- a. simplification de la structure logique de la base;
- b. amélioration de l'exhaustivité et de la cohérence de la couverture en texte intégral par:
  - exploitation des sources principales des textes en amont de l'OPOCE (en tenant compte des développements dans les Institutions) et en améliorant les outils d'exploitation des textes provenant de l'OPOCE;
  - extension de la couverture aux actes de gestion s'adressant aux PME, importateurs/exportateurs, douaniers, agriculteurs, etc.
  - introduction des textes des mesures nationales d'exécution des directives communautaires (si les ressources sont disponibles);
  - introduction des textes des actes préparatoires (documents COM, résolutions de la procédure législative, avis non-conformes et d'initiative du P.E. et du CES) dans la mesure où ils peuvent être récupérés sous format électronique;
  - introduction des saisines de la Cour de Justice et accélération de l'introduction des textes des arrêts et conclusions des avocats-général; introduction des affaires radiées;
  - introduction des textes des questions parlementaires et des réponses (projet INSIS ou récupération des JOs);
  - meilleur traitement des tableaux;

- c. accélération des mises à jour (p.ex. par transfert électronique des données du JO, par des outils d'analyse en ligne et par la rationalisation des procédures de gestion) et amélioration de la présentation des données (p.ex. par décomposition des textes en plusieurs champs comme PREAMB, CORPUS, ANNEXE etc.); il faudra arriver à un délai de m.à.j d'un jour pour la partie analytique (actuellement 2 semaines c.à.d. accélération de 86 %) et la partie textuelle (actuellement 4 semaines c.à.d. accélération de 93 %)
- d. possibilité de créer plus facilement les secteurs encore manquants;
- e. possibilité d'introduire ultérieurement des modules d'analyse morphologique et syntaxique pour améliorer l'indexation du texte libre dans les bases.
- f. diminution, voire élimination, de la captivité de l'application par rapport aux configurations et aux procédures utilisées actuellement tout en mettant en pratique les principes des systèmes ouverts.
- g. réduction des ressources utilisées actuellement pour la gestion et la maintenance du système.

## 11.2 Amélioration de l'accès au système

- a. rendre l'accès à CELEX et au droit communautaire le plus convivial possible par
  - la création d'interfaces conviviales;
  - l'implémentation du CCL;
  - l'assistance en ligne;
  - la fourniture de supports pédagogiques automatisés;
- b. mettre à la disposition des utilisateurs une série de services pratiques tels que :
  - la disponibilité d'une version coordonnée des actes modifiés;
  - la création d'un outil de commande en ligne des documents stockés dans l'archive optique à l'OPOCE;
  - la possibilité de télécharger les textes sélectionnés dans les systèmes de traitement de textes, en assurant l'introduction du texte intégral en typographie riche (N.B. l'introduction du riche dans CELEX est une condition nécessaire mais pas suffisante pour la mise en oeuvre d'un système performant de téléchargement en général).
  - la disponibilité d'un didacticiel sur PC pour l'apprentissage du système (Computer Aided Instruction System).

Dans la présente étude de faisabilité sont traités tous ces objectifs à l'exception des interfaces conviviaux, des supports pédagogiques automatisés, du téléchargement et du document delivery.

### III. SPECIFICATIONS UTILISATEUR

#### III.1 Consolidation des souhaits des utilisateurs finaux

Les utilisateurs de CELEX expriment leur souhait de voir le système amélioré depuis au moins 1985. Leur suggestions ont été exprimées en maintes reprises:

- En 1985 une étude de BUTLER-COX (réf. 3) sur les systèmes/bases de données juridiques en Europe n'as pas accordé de très bons points à CELEX en ce qui concerne la facilité d'accès et la disponibilité des versions coordonnées des textes législatifs modifiés.
- Pendant les deux réunions que tient par année le Groupe Informatique Juridique du Conseil, formule des nombreuses suggestions et propositions sur les besoins des spécialistes du droit communautaire et leur couverture par CELEX. Des demandes en ce sens ont été aussi formulées par la Commission Juridique et des Droits des Citoyens du Parlement Européen (ref. 4);
- Afin de mieux consolider les souhaits des utilisateurs finaux du système un questionnaire a été lancé en fin 1989 auprès des quelques 800 utilisateurs EUROBASES (en ce moment), 116 desquels ont répondu (réf. 5);
- Après chaque cours de formation CELEX, les participants remplissent une fiche d'évaluation sur laquelle ils marquent souvent leurs observations concernant la structure, la logique et les procédures d'accès à CELEX. Dans le cadre de la rédaction de la préanalyse et des études de faisabilité sur la modernisation de CELEX on a tenu compte aussi de ces suggestions provenant de plusieurs centaines d'utilisateurs. On n'a pas oublié non plus les problèmes signalés par les utilisateurs au HELPDESK
- Sur base des suggestions mentionnées supra et avant le lancement de la première etude de faisabilité concernant la Modernisation de CELEX (réf. 2) il a été jugé oppurtun de tester les options stratégiques retenues avec l'implementation d'un prototype de la nouvelle base de diffusion. Ce prototype, a été présenté au Groupe Informatique Juridique du Conseil et aux Informations Officers de la Commission; il a aussi été ouvert pendant une période limitée à une population reduit de quelques 20 utilisateurs externes (experimentés en CELEX et néophytes) qui ont formulé leurs observations par conséquent.
- Finalement dans le cadre de la redaction des spécifications utilisateur l'expérience acquise par les membres de l'équipe CELEX (introduction des données, administration, maintenance, formation, diffusion) ainsi que les résultats d'une analyse approfondie des données introduites dans CELEX ont été valorisés au maximum.

### III.2 Méthodologie

Comme au stade de la préanalyse, l'élaboration des caractéristiques du nouveau système comme elles sont perçues d'un point de vue de l'utilisateur suivra la méthodologie décrite ci-après. Cette méthode permet une meilleure compréhension du projet dans son ensemble et un énoncé concret de tous les problèmes identifiables à ce stade du projet.

- a. A partir de la situation existante, c'est à dire l'état actuel de la base de données et l'expérience acquise sur cette base par ceux qui l'utilisent et la font fonctionner, quelles doivent être les caractéristiques du nouveau système du point de vue de l'utilisateur final (c'est-à-dire les spécifications de la nouvelle base de diffusion).
- b. Sur base de ces spécifications, et après identification des sources existantes ou potentielles d'information brute à introduire dans le système, quelles sont les spécifications des programmes et procédures destinés à transformer ces données vers le produit final souhaité (c'est-à-dire les spécifications du système de production).
- c. En fonction des besoins des "applications client", (p.ex. production des profils, livraisons aux serveurs) quels éléments techniques sont à mettre en oeuvre pour permettre l'exploitation de CELEX en tant que fournisseur de produits de diffusion dans des conditions satisfaisantes du point de vue économique et fonctionnel.
- d. En tenant compte des différents points énoncés ci-dessus, quelles devraient être les spécifications des programmes et des procédures permettant d'assurer le suiti de l'exploitation courante dans les meilleures conditions d'économie et de transparence ("manager friendliness").

Pour chacun de ces ensembles de spécifications, il a été tenu compte de la possibilité éventuelle d'intégrer les développements correspondants dans le contexte du projet IDS. C'est la raison pour laquelle le "System Supplier" appartient à l'équipe IDS de l'unité IDA (Ingénierie des Applications).

### III.3 Bases de diffusion

#### a. Couverture

Analysé par secteur, la couverture du nouveau CELEX est reprise dans la Table 1 ci-après (les estimations du volume des textes intégraux à ajouter figurent en annexe III de l'étude de faisabilité TEXTERFACE, ref. 2).

Il faut noter ici que dans la préanalyse et dans l'étude de faisabilité TEXTERFACE, il a été proposé de substituer le secteur 7 "Mesures nationales d'exécution" par un secteur 7 "Version coordonnée des documents modifiés". Suite aux observations de plusieurs utilisateurs et à la demande des représentants des Etats Membres des groupes "Informatique Juridique" et "Affaires Générales" du Conseil, il a été décidé de ne pas procéder à cette substitution et de garder le secteur 7 "Mesures nationales d'exécution". Dans les documents de ce secteur seront introduits, si les ressources le permettent, les textes intégraux des mesures nationales qui correspondent à la version linguistique concernée. Les versions coordonnées des textes modifiés se trouveront dans leurs secteurs d'origine pourvues de numéros de documents distinctifs. Les estimations des textes coordonnés à saisir, figurent dans l'étude de faisabilité TEXTERFACE.

La systématisation de la "couverture différenciée" disparaîtra (pour une explication v. Annexe 1E point E de la Préanalyse). Chaque document figurera dans les bases avec son texte dans la mesure où celui-ci aura été traduit et sera disponible sur support magnétique.

Les documents dont le texte n'aura pas été traduit dans l'une des langues officielles figureront dans la version linguistique correspondante de CELEX sous forme de données bibliographiques, mais sans texte, avec une mention appropriée pour l'information des utilisateurs.

TABLE 1 : Couverture du CELEX modernisé

- Secteur 1 (Traités) :
  - \* texte intégral dans chaque langue;
- Secteur 2 (Relations extérieures) :
  - \* texte intégral;
  - \* plus de cohérence (en ce qui concerne les textes) entre les documents du secteur 2 et ceux du secteur 3 les adoptant;
- Secteur 3 (Droit dérivé) :
  - \* texte intégral dans les langues;
  - \* suppression des références aux propositions de modification après l'adoption de ces dernières;
  - \* introduction, dans les directives, de champ spécifique destiné à accueillir les références des mesures nationales d'exécution de ces directives (en complément au secteur 7);
  - \* introduction des actes de gestion (durée de validité de quelques semaines); ces actes seront écrasés après leur fin de validité pour être remplacés par des actes nouveaux;
  - \* introduction des rectificatifs comme documents indépendants en relation (mod/ult) avec le document initial;
  - \* introduction de la version coordonnée des actes modifiés;
- Secteur 4 (Droit complémentaire) :
  - \* texte intégral;
- Secteur 5 (Travaux préparatoires) :
  - \* accélération de l'introduction des documents;
  - \* introduction du texte des propositions pendantes devant le Conseil;
  - \* introduction du texte des avis du P.E. et du C.E.S.;
- Secteur 6 (Jurisprudence) :
  - \* extension de la couverture aux saisines, et radiations;
  - \* accélération de l'introduction des textes des arrêts et conclusions (par chargement dans les bases de chaque version linguistique dès la disponibilité de la traduction, et non plus après la publication au recueil);
- Secteur 7 (Mesures nationales d'exécution) :
  - \* introduction du texte intégral des mesures nationales d'exécution communiquées à la Commission. Pour les documents du secteur 7 les champs DISP.XX contiendront les références de publication des MNEs du pays correspondant et éventuellement le texte s'il est disponible pour les versions linguistiques correspondantes (p.ex. dans la base CLX les documents du secteur 7 contiendront les champs DISP.FR, DISP.BE, DISP.LU tandis que dans la base CEN ces mêmes documents contiendront les champs DISP.GB et DISP.IE).
- Secteur 8 (Décisions de jurisprudence nationale se référant au droit communautaire) :
  - \* ce secteur doit encore être créé, une étude de faisabilité spécifique a été lancée dans le cadre du programme INSIS;
- Secteur 9 (Questions parlementaires) :
  - \* introduction du texte des questions et des réponses, via une passerelle de transfert avec les documents correspondants dans les bases internes du PE ou de la Commission;
- Secteur 10 (Doctrines relative au droit communautaire) :
  - \* ce secteur doit encore être créé; la décision définitive de mise en place n'a pas encore été prise.

## b. Structure

La définition des nouvelles bases de diffusion est le résultat d'une analyse approfondie des données que contient CELEX. Sur base de cette analyse une rationalisation de la structure de la base était nécessaire. On a ainsi réduit le nombre des champs utiles à 62 (de 80) et le nombre des lexiques utiles à 7 (de 31).

Cette analyse a été effectuée sur base des critères suivants:

- indépendance aussi large que possible de la nouvelle structure du SGBD utilisé;
- regroupement des descripteurs communs dans des ensembles uniformisés en vue de l'utilisation dans tous les cas d'un vocabulaire bien contrôlé par une seule table de valeurs possibles; les champs qui contiennent du vocabulaire contrôlé du type bibliographique ont été regroupés dans le lexique BIB pour faciliter les recherches structurées;
- splittage des textes des documents en plusieurs "zones" qui correspondent à des champs séparés (p.ex. PREAMB, CORPUS, ANNEXE) afin de faciliter la visualisation des textes repérés. Tous les champs textuels ont été regroupés dans un seul lexique TXT; ce lexique déclaré "lexique implicite" permettra à l'utilisateur d'effectuer des recherches en langage presque naturel (ressemblant à un "feuilletement" des volumes des JOs et des RECUEILS);
- uniformisation des dates et regroupement des champs correspondants dans un seul lexique DATE;
- rationalisation de la présentation des relations entre documents; fusion des relations du même type dans un seul champ et regroupement des champs avec des relations dans un seul lexique REL;
- possibilité d'effectuer des recherches sur des descripteurs communs (expansion de la recherche ou sélection déduite);
- possibilité d'utilisation en synonymie des noms de champs CCL (ISO) pour faciliter l'interrogation par le module CCL;
- prévision de l'utilisation pour les recherches d'un thésaurus avec hiérarchie, synonymie et relation de groupe; ceci facilitera les recherches structurées, en combinaison avec l'attribution pour tous les documents de descripteurs EUROVOC ainsi que de codes de classification (correspondants aux matières traitées dans le cadre des travaux des Institutions); pour assurer un dénominateur commun la maintenance d'EUROVOC sera utilisé pour tenir compte des besoins de CELEX.

En plus il faut souligner les points suivants:

- la structure du numéro des documents CELEX a été rationalisée et il a été tenu compte de l'approche de l'an 2000; enfin l'introduction du numéro logique comme du numéro physique permettra une meilleure gestion des documents longs pour



- lesquels il faut créer des documents-suite (phénomène lié au SGBD utilisé);
- la nouvelle structure du champ AUTEUR permettra une série de recherches pour établir des statistiques comparatives pour les documents du Parlement européen;
  - l'introduction des documents indépendants pour les rectificatifs permettra leur saisie exhaustive et une meilleure gestion des liens entre les modifications ultérieures des documents;
  - l'introduction d'un champ DISP.NAT dans chaque directive du secteur 3 et la création des champs DISP.XX (XX = indicatif du pays) dans le secteur 7 permettront une recherche sur l'existence de mesures nationales d'exécution ainsi que la visualisation du texte de ces mesures si il a été communiqué à la Commission;
  - l'introduction d'un dispositif souple pour aider l'utilisateur en ligne (lexique MANUEL);
  - enfin l'introduction d'un champ OPOCE permettra le lancement du document delivery à partir du système d'archivage optique de l'OPOCE (ARCDON); l'utilisateur sera informé de l'existence de ce document dans ARCDON, ainsi que du prix à payer pour le recevoir.

La nouvelle structure des bases de diffusion CELEX est reprise en Annexe 2. La définition du contenu des champs est analytique. Il est aussi annexé à la présente E.F. un premier projet de Manuel d'utilisateur pour la version française du CELEX modernisé Annexe 3).

### c. Multilinguisme

Comme il est rappelé en introduction, CELEX doit être disponible dans chacune des neuf langues officielles des Communautés. En plus l'introduction de tous les caractères spéciaux dans les bases CELEX permettra une meilleure intégration du système dans l'environnement bureautique de la Commission en facilitant en même temps la vie des utilisateurs externes qui voudraient télécharger des textes législatifs

L'introduction de la typographie riche dans les bases de diffusion actuelles est déjà mise en route dans le cadre du sous-projet 1: TEXTERFACE; à partir du 13/3/1990 tous les nouveaux textes qui entrent dans les bases sont en "riche" et une opération de rechargement de textes anciens (textes IC2) est prévue.

Les codes utilisés respectent les normes internationales en vigueur, à savoir les normes ISO 8859 (/1 pour les langues latines et /7 pour le grec). L'expérience acquise dans le cadre de la création de la version grecque de CELEX était très utile en la matière. La solution adoptée ne permet pas la coexistence au sein d'un même texte de caractères grecs et "latins spéciaux" (ce qui nécessiterait un environnement 16 bits ou le recours à un système compliqué de séquences d'échappement), mais constitue un excellent

compromis pratique: outre le fait qu'une telle situation de coexistence est relativement rare, elle ne sert jamais de base à la recherche. Dans la pratique on procède à une translittération "intelligente" de ces caractères, comme c'est le cas actuellement pour les symboles mathématiques.

Dans le système actuel il est assez fréquent que les utilisateurs (notamment à l'intérieur des Institutions) passent d'une version linguistique à une autre au cours d'une même recherche; ceci est faisable grâce à la possibilité de pouvoir sauvegarder et réexécuter une stratégie de recherche.

#### d. Tableaux et graphiques

La restitution sur un écran 80 x 24 ou 132 x 24 de tableaux et graphiques figurant dans les textes publiés au J.O. demeure un problème majeur. Pour que ces éléments puissent être introduits dans les bases de façon à en rendre la visualisation possible, une normalisation des formats et des nouvelles possibilités du logiciel documentaire de traiter et mettre en forme ces données, sont nécessaires. A l'heure actuelle, cette mise en forme est réalisée - dans la mesure du possible et pour les tableaux de taille réduite - lors du seul encodage des textes non fournis par l'OPOCE.

#### e. SGDB pour les bases de diffusion

Dans le cadre de la Préanalyse une analyse comparative a été effectuée entre deux SGBDs en ce qui concerne les fonctionnalités indispensables et les performances requises pour les bases de diffusion de CELEX. Il est vrai que les fonctionnalités jugées à ce moment indispensables n'étaient couvertes que par le progiciel MISTRAL, mais il ne faut pas oublier que le seul progiciel mis en concurrence était BASIS. Il paraît par conséquent possible qu'à l'avenir d'autres progiciels, basés sur des modèles plus modernes que le modèle booléen utilisé actuellement, rempliront encore mieux les exigences de CELEX.

Les fonctionnalités exigées du SGBD qui sera utilisé pour les bases de diffusion de CELEX sont reprises en Annexe 4 sous la forme d'une liste de contrôle; dans le cadre d'une mise en concurrence éventuelle il faut aussi tenir compte des performances sous des conditions réelles d'exploitation.

### III.4 Le système de Production

Bien qu'indirectement perçues par les utilisateurs, l'alimentation et l'administration des bases de diffusion méritent une attention particulière car leur efficacité conditionne la qualité, la fiabilité et l'actualité de l'application.

Les procédures actuelles d'introduction des données dans CELEX sont compliquées, mais le concept même d'un "fichier maître neutre" (appelé ARCHIVE dans le jargon CELEX) pour la production de la partie analytique des documents ne peut pas être mis en question. Ce concept correspond en plus parfaitement aux nouveaux principes de la Commission en matière de développement de systèmes d'information. Les avantages offerts par un fichier maître central ne sont plus à démontrer (production de sept - et bientôt neuf - versions linguistiques à partir d'un noyau unique, contrôle facile de la cohérence, réduction du risque de captivité de l'application, etc.) et le principe lui-même sera donc retenu.

Le système de production des bases CELEX doit assurer les fonctions suivantes:

- Fonction Analyste
- Fonction Texterface
- Fonction Data Base Manager (DBM)
- Fonction Data Base Administrator (DBA)
- Fonction Maintenance

L'architecture générale de l'application CELEX adoptée dans le cadre de la préanalyse est de type distribué; il prévoit:

- des bases de diffusion qui seront toujours sur machine centrale;
- une base de production des textes sur machine locale;
- une base de production de la partie analytique (nouvelle ARCHIVE);
- l'alimentation/saisie et la gestion de cette base de production seront sur une machine locale.

Puisque la nouvelle ARCHIVE et les bases de diffusion (BDD) seront sur des machines différentes, les fonctions DBM, DBA et Maintenance auront une partie locale et une partie centrale.

Les caractéristiques principales du système de production requis par les utilisateurs du système, à savoir les membres de l'équipe CELEX, sont repris ci-dessous:

- possibilité d'analyse en ligne avec reprise dans la nouvelle ARCHIVE des informations déjà encodées sur les bandes FORMEX et avec validation automatique des descripteurs du vocabulaire contrôlé par tables; introduction du code abrégé avec affichage de l'équivalent français; possibilité d'affichage des valeurs possibles autour d'un descripteur probable;

- génération automatique de certaines données (p.ex. Secteur 7);
- introduction des nouveaux documents; modification des documents existants (ad hoc et corrections systématiques);
- activation des contrôles automatiques sur base du NUMERO du document qui est validé manuellement; sur base du NUMERO activation des écrans de saisie spécifiques par document;
- introduction des nouveaux descripteurs du vocabulaire contrôlé dans toutes les langues; édition par langue(s) et par classe(s) de descripteurs; possibilité de corrections systématiques du thesaurus (répercutées sur les bases de diffusion);
- gestion des niveaux de secrets des documents isolés, pour des classes de types de documents et pour des champs;
- gestion des champs textuels qui sont différents pour chaque version linguistique (p.ex. DISP.NAT, TRIB.NAT);
- gestion automatique des relations entre documents;
- gestion des correspondances entre les rubriques de la base locale et les champs des bases de diffusion;
- répartition du travail d'analyse (contrôle de l'accès concurrentiel aux documents);
- interconnexion avec le système d'introduction/gestion de la partie textuelle (TEXTERFACE) avec intégration totale dans la phase finale de mise en production; possibilité de correction des textes; possibilité d'encodage des textes nouveaux;
- fonctions DBM, DBA et de maintenance.

Plus particulièrement les fonctionnalités requises par chaque tâche sont les suivantes:

a. Fonction Analyste

- Accès sur le même écran (à intégrer dans le cadre d'un futur interface graphique):
  - \* à la nouvelle ARCHIVE pour introduire/valider des données, valider des documents, introduire des documents nouveaux;
  - \* aux Bases de Diffusion (BDD) pour contrôle et/ou consultation des données historiques;
  - \* au texte français du document communiqué par l'OPOCE ou le SG, pour les documents qui sont en cours d'analyse;
- Validation des données introduites automatiquement des bandes FORMEX; introduction des codes dans les autres rubriques analytiques; validation finale du document avant transfert vers les BDDs.

## b. Fonction Texterface

- Accès sur le même écran:
  - \* à l'ARCHIVE nouvelle pour introduire/valider des données textuelles;
  - \* aux Bases de Diffusion (BDD) pour contrôle, pour consultation des données historiques et afin de sélectionner des ensembles de documents pour les transférer en local afin de les traiter (production des versions consolidées, corrections systématiques de textes);
  - \* aux textes multilingues des documents communiqués par l'OPOCE ou le SG, pour les documents qui sont en cours d'analyse;
  - \* au texte de tout autre document qu'il a encodé sur le système local ou transféré du système central.
- Validation du splittage en champs textuels et des balises SGML vis PG, MORE, VI ou Q-one; possibilité de consulter un listing avec les débuts et fins du splittage;
- sur base des mots nouveaux introduits dans les BDDs, correction éventuelle des textes [Pour un document nouveau A: le jour x, le A est chargé dans MISTRAL, on détecte les mots nouveaux (fautes), on corrige sur le local le jour x+i, on renvoie le document à MISTRAL le x+i et on fait une élimination des fréquences 0 (i=? pour ne pas avoir des fautes dans la base)];
- possibilité d'encoder des textes nouveaux (documents vraiment nouveaux, documents provenant de fusion des documents existants - codification - sur le central), directement ou par association des fichiers Q-office).
- possibilité d'effectuer des corrections/ajouts systématiques dans les titres (paragraphe 2TI actuels; fusion avec champs ORACLE).

## c. Fonction Data Base Manager (DBM)

- Accès aux fonctions Analyste et Texterface pour toutes les langues;
- possibilité de valider en supplémentaire des analyses de certains analystes et de superviser leur travail (p.ex. analystes nouveaux);
- possibilité d'annuler des documents ou des relations entre documents;
- possibilité de mettre à jour les tables multilingues de CELEX et de gérer le thésaurus;

#### d. Fonction Data Base Administrator (DBA)

- Accès à l'ARCHIVE au plus haut niveau;
- récupération des données de l'OPOCE, du PE et de la CJUS, puis lancement de leur introduction dans le circuit d'analyse et de traitement linguistique; introduction des données disponibles sur bandes et/ou en provenance d'autres sites, gestion des textes trop longs; gestion du stock des textes conservés en local;
- répartition des documents à valider (manuellement par la diffusion d'un listing avec numéros chaque jour ou par courrier électronique) aux analystes/linguistes;
- extraction de documents nouveaux et modifiés (partie analytique et texte) de l'ARCHIVE et du TEXTERFACE; lancement de la génération des traductions automatiques dans toutes les langues, et transfert des données sur le site central avec lancement des m.à.j des BDDs;
- Outre les outils dont le DBA doit disposer sur le site local, il doit disposer aussi certains outils de gestion des bases sur le site central (outils pour effectuer des réorganisations, pour éliminer des termes à fréquence 0, pour modifier/ajouter des éditions cataloguées, des profils, des rapports et des macros, pour modifier/ajouter des messages d'accueil avec des informations utiles pour les utilisateurs);'

#### e. Fonction Maintenance

La mise en oeuvre d'un système aussi complexe que le CELEX modernisé est en soi une opération d'engrevure; il serait illusoire de considérer que le système fonctionnera d'une façon idéale et correcte surtout pendant la période initiale. En plus le système CELEX sera utilisé par des humains intelligents dont les besoins sont sensés évoluer dans le futur. Par conséquent, il faut prévoir aussi une fonction maintenance avec un double axe: maintenance corrective et maintenance évolutive.

Pour un système bien conçu et construit, la charge de la maintenance corrective, importante au début, doit diminuer avec le temps; par contre la maintenance évolutive se verra augmenter avec le temps jusqu'au prochain plan de modernisation du système.

Les travaux qui seront demandés dans le cadre de la maintenance évolutive seront en principe les suivants:

- introduction d'une nouvelle classe de type de documents pour laquelle il faudra prévoir des contrôles, des validations et des traitements spécifiques;
- introduction d'un nouveau champ (textuel ou analytique) dans les BDD surtout dans le cas spécifique où le nouveau champ des BDD provient de la fusion des champs existants; dans ce cas il faudra mettre à jour la table de correspondances balises locales-champs sur central;

- introduction des nouvelles versions linguistiques; bien que cette fonction ne soit pas de première priorité après la construction des bases espagnole et portugaise, il ne faut pas oublier de prévoir, pour le cas échéant, l'introduction d'autres versions linguistiques.
  
- possibilité de modifier/ajouter/changer-le-nom/ajouter-nom-synonyme/regrouper-dans-lexique, pour un champ des BDD.

#### IV. STRUCTURE DU SYSTÈME D'INFORMATION : SOLUTION PROPOSÉE ET SOLUTIONS ALTERNATIVES

##### IV.1 Contexte général du système :

Comme il a été déjà mentionné (§ III.3 Le système de production) l'architecture générale de l'application CELEX, adopté dans le cadre de la préanalyse, est de type distribué: CELEX peut être considéré comme étant composé de trois sous-systèmes (modules) :

- le système des bases de diffusion (BDDs-CELEX);
- le système de production de la partie analytique des documents (ALICE-A avec ALICE = ALimentation CElex);
- le système d'alimentation de la partie textuelle (ALICE-T dit TEXTERFACE) (v. ref. 2).

##### a. Le système des bases de diffusion (système BDDs-CELEX)

La fonction du système des bases de diffusion (BDD) est de stocker les unités documentaires CELEX complètes (avec leur partie analytique et textuelle) pour la diffusion publique de l'information juridique Communautaire. Le système BDDs-CELEX est un vrai système de diffusion de l'information, organisé comme un ensemble de bases de diffusion (une base par version linguistique).

Le schéma IV.1 présente une vision globale du flux d'informations du système BDDs-CELEX avec ses entrées et ses sorties ainsi que ses relations avec d'autres systèmes.

##### Entrées :

\* En provenance du système ALICE-A :

- Par version linguistique, la partie analytique de l'unité documentaire CELEX dans un format standard (suivant la syntaxe SGML).
- Par version linguistique, les termes du langage documentaire (thésaurus) pour la consultation des bases et dans un format standard (suivant la syntaxe SGML).

\* En provenance du système ALICE-T :

- Par version linguistique, les textes des documents faisant partie de la couverture CELEX (JO L et C, Recueil, documents COM, Questions Parlementaires, encodage versions codifiées et textes des Mesures Nationales d'exécution).

\* Les requêtes/demandes des utilisateurs pour obtenir de l'information :



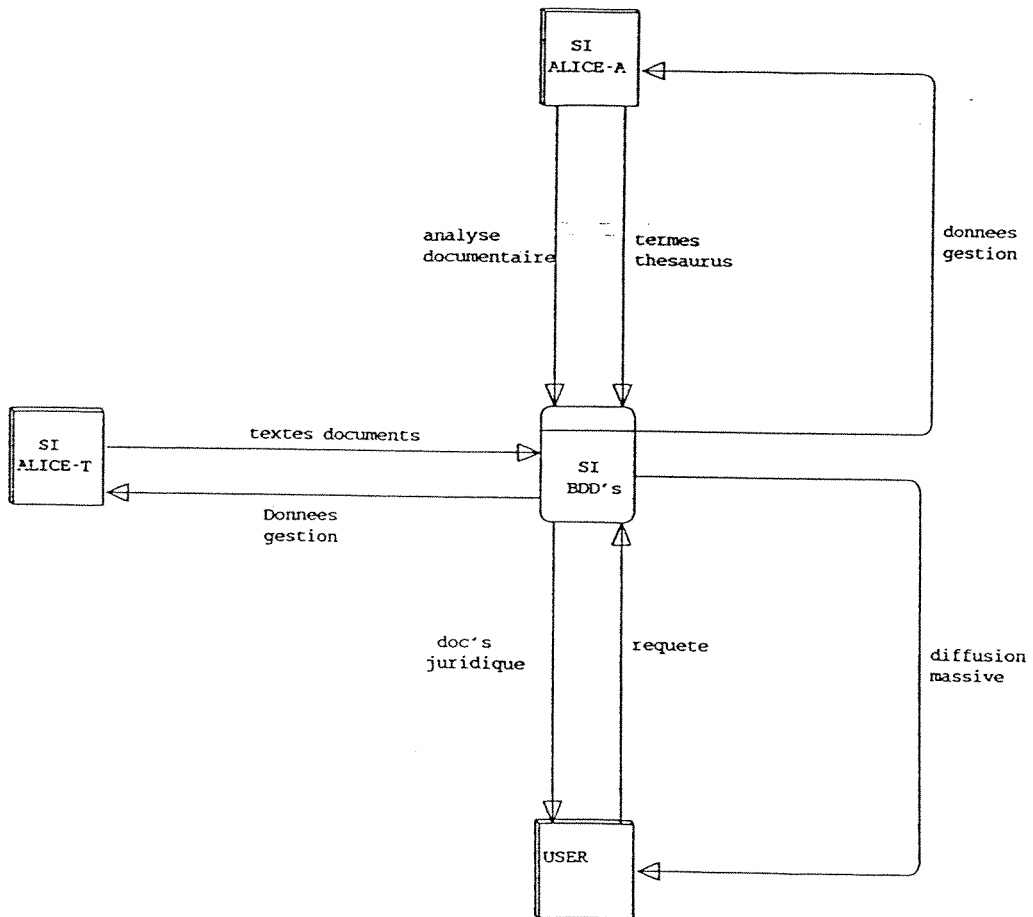


Fig. IV.1 DIAGRAMME DE CONTEXTE (ISAD)

BASES DE DIFFUSION (BDD's)

Sorties :

- \* L'information juridique répondant aux requêtes des utilisateurs.
- \* La diffusion sélective de grands lots d'informations (répertoire, bandes serveurs, profils).
- \* Information sur le processus de gestion envoyée vers les sous-systèmes ALICE-A et ALICE-T.

Le système fournira également de l'information de gestion (statistiques, compte rendu d'exécution) de gestion sur son propre fonctionnement et sur la cohérence avec les autres sous-systèmes (ALICE-T et ALICE-A).

b. Le système de production de la partie analytique (système ALICE-A)

La fonction de ce sous-système est de "produire" et alimenter l'information analytique associée aux documents qui font partie de la couverture CELEX.

Le schéma IV.2 présente une vision globale du flux d'informations du système ALICE-A avec ses entrées et ses sorties ainsi que ses relations avec d'autres systèmes.

Entrées :

- \* **Le Parlement Européen** : pour l'analyse documentaire des résolutions et autres actes préparatoires du Parlement (secteur 5) et les questions/réponses parlementaires (secteur 9).
- \* **La Cour de Justice** : pour l'analyse documentaire des textes des jurisprudences communautaires (secteur 6) et nationales (secteur 8).
- \* **Le Conseil** : pour l'analyse des actes conventionnels du Conseil faisant partie de la couverture CELEX (accords et conventions avec les pays tiers) (Secteur 2).
- \* **Le Comité Economique et Social** : pour l'analyse des avis et autres documents préparatoires du CES faisant partie de la couverture CELEX (Secteur 5).
- \* **Eventuellement autres services de la Commission (SG, DG's)** : pour l'analyse de certains actes de leur domaine de responsabilité qui font partie de la couverture CELEX.
- \* **L'équipe des analystes CELEX** : pour l'analyse documentaire des autres documents de la couverture (actes législatifs, actes préparatoires de la Commission, ...) plus la gestion des termes du langage documentaire. L'équipe CELEX effectue également la vérification et la correction éventuelle des analyses documentaires en provenance d'autres institutions.

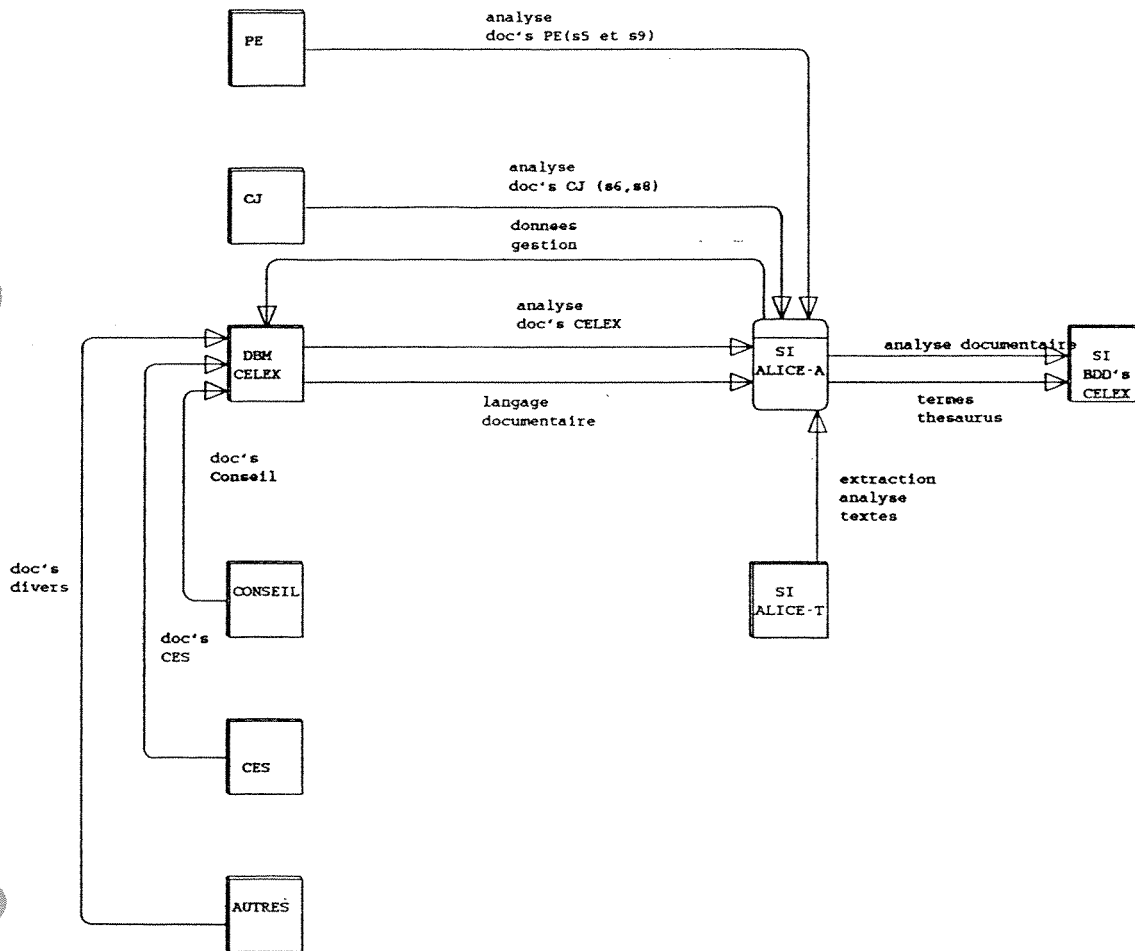


Fig. IV.2 DIAGRAMME DE CONTEXTE (ISAD)

ALICE-A

Remarques :

- A l'heure actuelle les analyses documentaires des actes CES, CONSEIL et d'autres services de la Commission transitent par l'équipe de gestion CELEX sous forme papier.
- Il serait possible de produire une première esquisse d'analyse à partir du sous-système ALICE-T; cette esquisse peut être obtenue pendant la phase d'analyse automatique des textes intégraux lors du chargement de ces textes à partir des JOs.

Sorties :

- \* Pour chaque version linguistique, la partie analytique de l'unité documentaire CELEX.
- \* Pour chaque version linguistique, les termes du langage documentaire pour la gestion du thésaurus dans le base de diffusion.
- \* Information de gestion destinée au DBM.

c. Le système d'alimentation de la partie textuelle (système ALICE-T)

La fonction du système ALICE-T est de gérer et de traiter les différentes sources d'alimentation des textes de CELEX ainsi que de produire la partie textuelle de l'unité documentaire CELEX destinée à être chargée dans les bases de diffusion CELEX.

Le schéma fig. IV.3 présente une vision globale de ce système avec ses différentes entrées (sources d'alimentation CELEX) et les sorties produites par ce système.

Entrées :

- \* L'OPOCE pour l'alimentation des textes publiés dans le Journal Officiel (séries L,C).
- \* La Cour de Justice pour l'alimentation des textes publiés dans le Recueil de la Jurisprudence.
- \* L'équipe CELEX pour l'encodage externe et la saisie directe des textes.
- \* Le Parlement Européen pour l'alimentation des avis et autres documents du Parlement ainsi que des textes des questions/réponses parlementaires.
- \* Autres services de la Commission (SG, DG's) pour l'alimentation des divers types de documents qui font partie de la couverture CELEX, notamment les documents COM finaux.

Sorties :

- \* Pour chaque version linguistique, les textes des documents destinés à être chargés dans les BDDs; ces textes sont décomposés en champs selon une logique spécifique à CELEX et se présentent sous un format compatible MISTRAL produit à l'aide

d'un format intermédiaire basé sur SGML.

\* Certaines informations de type analytique qui vont servir pour l'initialisation d'une esquisse de la partie analytique de l'unité documentaire (alimentation du système ALICE-A).

\* Des informations de gestion destinées au DBM CELEX.

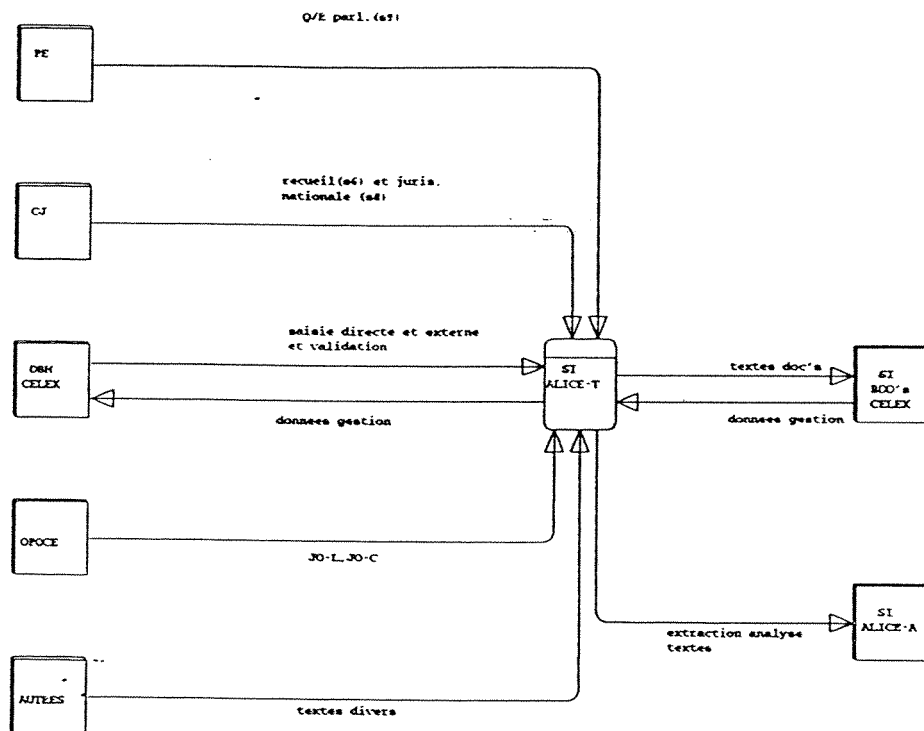


Fig. IV.3 DIAGRAMME DE CONTEXTE (ISAD)

ALICE-T (TEXTERFACE)

## IV.2 Architecture de la solution proposée : description fonctionnelle

### a. Généralités

L'élaboration de l'architecture du système d'information CELEX telle qu'elle se présente dans le cadre de la solution proposée, s'est basée sur les principes suivants :

- Architecture générale de type distribué avec application des concepts de base de production et base de diffusion.
- Le noyau du sous-système de production de la partie analytique (ALICE-A) sera une base de production appelée ARCHIVE (BDP-Archive) à caractère relationnel installée dans une unité locale accessible par les membres de l'équipe CELEX ainsi que par les institutions responsables de l'alimentation des secteurs de CELEX.
- L'accès à la BDP-Archive sera interactif avec la possibilité de procéder à l'analyse documentaire en ligne.
- La BDP-Archive est et restera une base unique et constitue le "moteur" de la production de la partie analytique de n versions linguistiques. Il faut se rappeler que ce concept est implémenté et fonctionne dans le système CELEX existant.
- Le sous-système ALICE-A fera partie de la même unité locale que le sous-système ALICE-T (ref. 2).
- Le sous-système de bases de diffusion (BDDs) sera organisé autour de 9 bases de diffusion (une par version linguistique)<sup>(1)</sup> accessibles tant par des utilisateurs internes qu'externes. Ces bases sont de type documentaire avec des fonctionnalités de recherche du texte intégral. Ces bases de diffusion sont et seront installées sur une unité centrale.

Le schéma IV.3 présente, sur la base de ces principes, le schéma de l'architecture de la solution proposée avec ses différents modules.

Par la suite nous allons décrire les différents modules ainsi que leurs caractéristiques les plus importantes. Les détails seront donnés dans les différentes annexes.

### b. Base de production de la partie analytique (BDP-Archive)

- La BDP-Archive aura comme fonction le stockage permanent et dans une forme codifiée de la partie analytique des unités documentaires CELEX qui sont destinées à être chargées dans les bases de diffusion.

(1) Le choix de cette organisation n'est pas dû seulement à des raisons historiques (système actuel) ou techniques (non disponibilité d'un système de gestion des bases de données documentaires multilingues) mais aussi au fait de viser plusieurs communautés linguistiques.

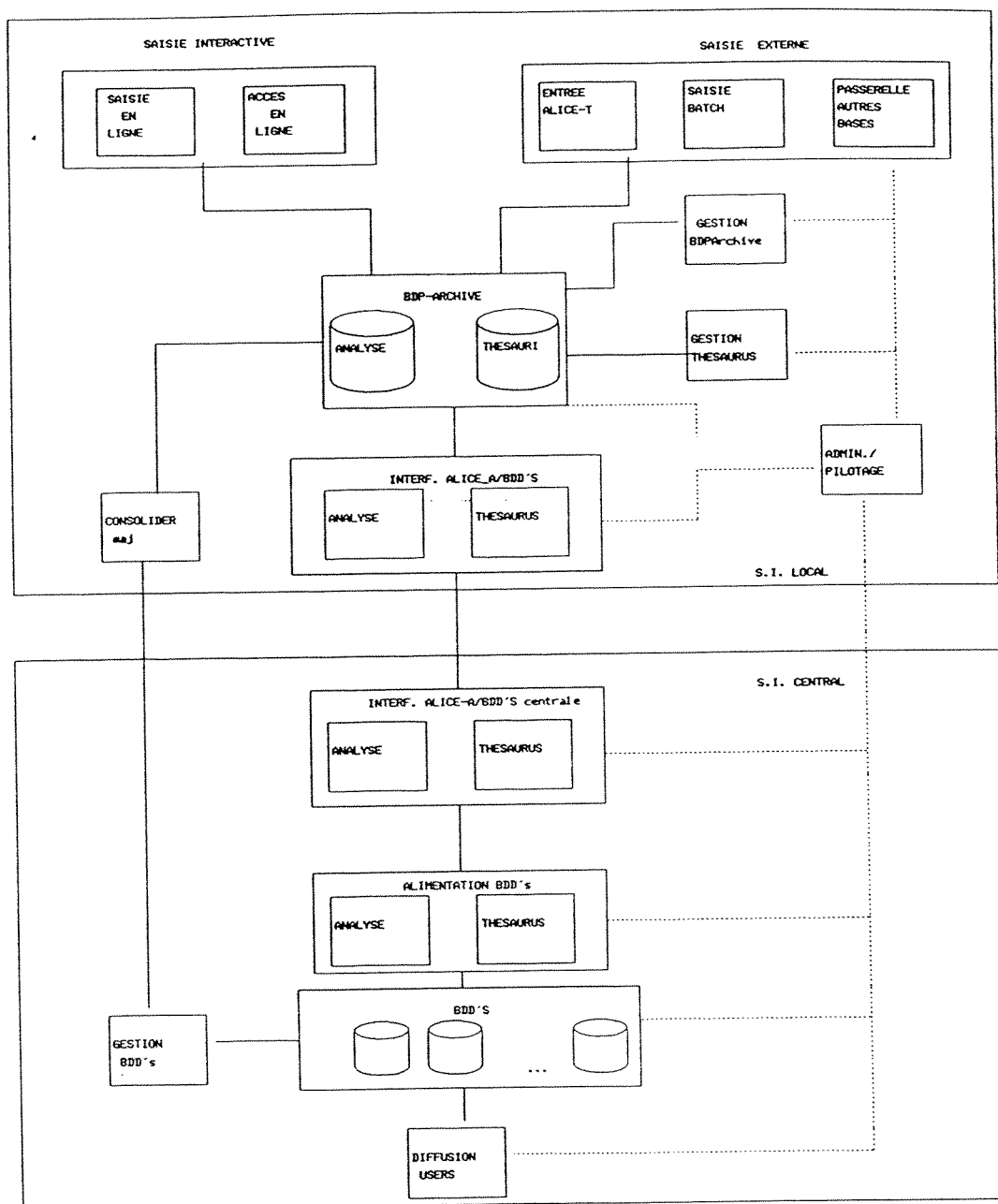


Fig. IV.4 ARCHITECTURE GENERALE SYSTEMES D'INFORMATION ALICE-A/ALICE-T/BDDs-CELEX

- Elle constituera donc le noyau essentiel pour la production des champs analytiques des BDDs CELEX.
- Il existe une relation d'injection entre les documents de la BDP-Archive et les documents dans la BDD. Le système à mettre en place garantira à tout moment cette cohérence.
- La BDP-Archive est orientée à la production des données et non vers la recherche documentaire. Par conséquent l'accès à l'information qui s'y trouve se fera :
  - \* à l'aide de menus formatés qui utiliseront les champs-clés des tables relationnelles définies ;
  - \* plus rarement à l'aide du langage d'interrogation SQL.
- La structure de cette base sera relationnelle et composée des groupes de tables suivantes:
  - \* les tables associées aux documents, avec la partie analytique d'un document CELEX dans une forme codifiée et indépendante de toute version linguistique ;
  - \* les tables de validation et de production des versions linguistiques qui contiendront les termes codifiés employés dans l'analyse documentaire d'un document CELEX et leur équivalence dans les différentes langues. Les tables seront organisées sous forme de thésaurus simple.
  - \* les tables de relations entre les documents (références croisées).

La définition détaillée de ces tables se trouve en Annexe 5.

### c. Alimentation/saisie des données analytiques

- Ce module comprendra toutes les fonctions (saisie et analyse en ligne, saisie complémentaire, passerelle avec d'autres bases, récupération de l'analyse des textes depuis le système ALICE-T, alimentation "batch") relatives à l'alimentation de la base BDP-Archive.
- Il sera organisé en deux sous-modules:
  - \* Sous-module de saisie directe et interactive avec les fonctions d'analyse en ligne et les fonctions de recherche dans la base. Toutes ces fonctions seront groupées sous un interface utilisateur unique.
  - \* Sous-module de saisie "batch". Ces fonctions seront intégrées dans l'interface spécifique au DBA.
- Pour l'alimentation "batch" on utilisera un format orienté document décrit à l'aide de la syntaxe SGML. Ce format est une extension de celui qui est utilisé pour l'alimentation de textes et doit donner la possibilité d'alimenter simultanément la partie analytique et la partie textuelle de l'unité documentaire CELEX.
- Le processus d'alimentation "batch" sera utilisé par la CJ et le PE; en cas d'erreur, ces Institutions auront la possibilité soit d'accéder au module interactif de saisie pour corriger,



soit de renvoyer à nouveau le document.

- La saisie de la partie analytique des documents se fera directement en ligne avec un ensemble d'écrans formatés selon le type de document à saisir/analyser. Ces écrans seront développés à l'aide du module de saisie du logiciel choisi pour implémenter la BDP-Archive et ils auront un même style de présentation.
- La plupart des champs analytiques seront complétés et validés interactivement, mais il y aura d'autres sources de données (passerelle avec d'autres bases, bandes avec des données analytiques, esquisse d'analyse en provenance de ALICE-T) pour lesquelles le contenu de certains champs pourra s'obtenir automatiquement. Dans ce cas les documents seront chargés provisoirement dans la base et, ensuite, l'équipe des analystes de CELEX procédera à la saisie complémentaire et la validation de tout le document.
- Comme la structure du document de la BDP-Archive est différente de la structure du document des BDDs, l'utilisateur aura toujours la possibilité de visualiser la traduction associée à un code; il disposera aussi d'une fonctionnalité qui lui fournira une présentation du document tel qu'il sera visualisé ultérieurement dans la BDD, après la mise à jour de celle-ci.

#### d. Gestion et administration de la BDP-Archive

- Ce module comprend toutes les fonctions de gestion et administration typiques d'une base de données : procédures de sauvegarde/restauration, définition et contrôle d'accès des utilisateurs (analystes et autres membres de l'équipe CELEX), gestion de la cohérence et de l'intégrité de la base, statistiques diverses, tableaux de bord sur le fonctionnement du système, chaîne INFOCELEX, distribution des documents non validés à analyser, ...
- Toutes les fonctions sont groupées dans une interface spécifique pour le DBA/DBM et leur mise en oeuvre se fera via le choix d'une option dans un menu.

#### e. Gestion thésaurus

- Il s'agit d'un module interactif pour la saisie (création et mise à jour) des termes repris dans les tables de validation (termes codifiés) et aussi des termes repris dans les tables avec les équivalences linguistiques.
- La définition de la structure du thésaurus est intégrée dans la BDP-Archive. Cependant, du fait de leurs caractéristiques spéciales, les fonctions de mise à jour auront une interface spécifique séparée.

#### f. Interface locale ALICE-A/BDDs-CELEX

- La fonction de ce module est de générer le format de chargement pour chacune des versions linguistiques des bases de diffusion via l'utilisation d'un format intermédiaire.

- Ce module est composé de deux parties :
  - \* L'interface pour la partie analytique qui transformera le document codifié stocké dans la BDPA-Archive en un document CELEX pour chacune des versions linguistiques.
  - \* L'interface thesaurus qui transformera les termes présents dans l'Archive sous forme codifiée en leurs equivalences dans les diverses versions linguistiques et les rendra prêts à être chargés sous forme de relation du type thesaurus dans chaque base de diffusion.
- Dans un souci de rendre le système indépendant du SGBD utilisé pour les BDDs, il sera utilisé un format intermédiaire décrit à l'aide de la syntaxe SGML qui, ensuite, sera converti dans le format final du SGBD des BDDs.
- Le processus de production de la partie analytique aura quatre phases :
  1. sélection dans la BDP-Archive des documents à transférer
  2. génération des unités documentaires pour chaque version linguistique dans un format intermédiaire décrit à l'aide de la syntaxe SGML
  3. conversion du format SGML dans le format d'entrée du logiciel de la base de diffusion
  4. transfert vers le système central à l'aide de "file transfert" MFTS.
- Le processus pour la production des relations "thesaurus" sera analogue.
- Le processus de production des versions linguistiques sera administré ou plutôt "pilote" par une table qui définira la correspondance entre le nom des champs dans la BDP-Archive et le nom des champs pour chaque version linguistique des BDDs.
- Dans le cas où le texte d'un document sera disponible avant le chargement en central de la partie analytique de ce même document, on pourra envisager la fusion des parties "analytique" et "textuelle" pour constituer un seul fichier de mise à jour (cohérence entre documents nouveaux et documents modifiés).

#### g. Structure des bases de diffusion BDDs-CELEX

- La fonction des BDDs est de stocker les unités documentaires CELEX (partie analytique et textuelle) pour la diffusion publique de l'information juridique.
- Les BDD's auront les caractéristiques suivantes :
  - \* pour chaque version linguistique il y aura une base de diffusion spécifique (pour l'instant il y a en tout 9 bases) ;
  - \* chaque base de diffusion sera une base de type documentaire à texte intégral interrogeable à l'aide d'un SGBD du type documentaire.

- La structure détaillée des BDDs CELEX se trouve en annexe 3; dans l'Annexe 4 est repris un projet du Manuel Utilisateur.
- Dans la définition des BDD's sont à signaler les aspects suivants :
  - \* une simplification importante de la partie analytique par rapport à la définition actuelle de la base avec élimination de certains traitements de gestion complexes (par exemple les rectificatifs).
  - \* chaque champ est défini de façon formelle et structurelle à partir de ses composants. Cette définition a été validée à l'aide d'une analyse syntaxique SGML (voir la définition SGML en Annexe 5),
  - \* la définition de la BDD sera indépendante de tout logiciel documentaire, et elle pourrait être implémentée sous un progiciel autre que MISTRAL,
  - \* traitement des documents-suite : étant donné qu'une restriction de progiciel MISTRAL consiste à avoir une taille maximale de document et une taille maximale de champ (ce qui entraîne la définition de documents-suite et de champs-suite), on a introduit le concept de numéro logique (NUMERO) et de numéro physique (REFER) pour permettre une meilleure gestion des documents-suite,
  - \* la création d'un système d'aide en ligne.

#### h. Interface ALICE-A/BDD's centrale

- La fonction de ce module est de gérer la réception des fichiers de chargement envoyés par la machine locale et de préparer les fichiers pour leur chargement dans les bases.
- Comme c'est déjà le cas pour l'interface locale, ce module sera composé de deux parties : la partie analytique et la partie thesaurus.

#### i. Alimentation des bases de diffusion

- Bien que le système de production sera unique, le module interface produira des fichiers à charger séparément dans chacune des versions linguistiques. L'objectif de ce module est d'organiser la mise à jour "parallèle" des BDD's tant de leur partie "analytique" que de leur partie "thesaurus".

#### j. Gestion des bases de diffusion

- Il s'agira des fonctions typiques de gestion d'une base de données (sauvegarde/restauration, gestion de cohérence,...) mais dans le cas de CELEX il faut tenir compte du fait qu'il y a 9 bases et qu'il faut faire une gestion en parallèle et, dans la mesure du possible, automatique (automatisation de l'environnement de gestion).
- Dans ce module se trouvent aussi :
  - \* des outils pour l'obtention des statistiques diverses (mise à jour, consultation) ;
  - \* un contrôle des résultats de la mise à jour des bases.

La plupart de ces outils ainsi que d'autres (par exemple la gestion d'une bibliothèque des requêtes, les éditions des descripteurs, la modification des éditions standardisées etc.) feront partie des outils du SGBD utilisés pour les BDDs.

#### k. Diffusion de l'information

- Ce module sera appelé à gérer la diffusion de l'information contenue dans les bases CELEX parmi un vaste ensemble d'utilisateurs et en utilisant différents mécanismes et supports :
  - \* accès interactif en langage natif MISTRAL ou CCL;
  - \* accès via une interface conviviale (UFI);
  - \* diffusion régulière sur support papier auprès des utilisateurs internes (profils);
  - \* production du Répertoire de la Législation Communautaire en vigueur dans toutes les langues;
  - \* diffusion sur supports magnétiques (gestion des serveurs externes);
  - \* mise en place d'un outil de commande en ligne interconnecté avec l'OPOCE pour l'envoi des versions papier des JOs publiés (document delivery);
  - \* gestion des accès utilisateurs;
  - \* éventuellement des outils destinés à la formation des utilisateurs finaux.

#### l. Administration et pilotage

- Le système d'information CELEX contiendra un ensemble d'éléments qui assureront l'administration ou plutôt le "pilotage" et guideront le processus d'alimentation et préparation de la génération des versions linguistiques (tables de contrôle et validation, tables de traduction, tables de correspondance entre champs, ...)\*. Ces éléments seront paramétrés de façon à permettre la préparation de la création de nouvelles versions linguistiques, nouveaux types de documents, nouveaux secteurs, ...
- Le but de ce module de pilotage/administration est de mettre à disposition du DBM toutes ces fonctionnalités sous la forme d'une interface uniforme, cohérente et automatisée. Évidemment ce module ne constituera en aucun cas un système de génération des applications ou des bases de données.
- Avec ces outils le DBM pourra gérer lui-même une partie de l'évolution des bases CELEX, étant donné que certaines actions, qui actuellement demandent l'intervention des développeurs, deviendront des actions de gestion. Parmi ces actions :
  - \* la création d'un nouveau type/classe de document;
  - \* la création d'un nouveau secteur;
  - \* la préparation d'une nouvelle version linguistique;
  - \* la création et modification des tables de validation et de traduction;

\* ces fonctions sont assurées aujourd'hui via des tables gérées avec THELEM (Table des Secrets, COPARCHAMP, TAB.CLX etc.)

- Ce module aura aussi une partie destinée aux développeurs pour diminuer l'effort de maintenance à fournir (définition d'une nouvelle base, création/modification d'un nouveau champ, modification de la structure d'un champ dans les bases de diffusion).

#### m. Outils de migration

Les deux systèmes, nouveau et ancien ont des différences structurelles importantes; par conséquent il faut élaborer des outils d'aide à la migration.

Ces outils couvrent :

- La migration de l'archive ancienne vers la BDP-Archive;
- La migration des anciennes bases de diffusion vers les nouvelles BDDs;
- La validation/vérification des résultats des processus de migration

#### n. Relation entre BDD-CELEX et BDP-Archive: gestion de la cohérence

- La cohérence entre la BDP-Archive et les BDDs sera gérée de façon à garantir que le nombre de documents soit le même et que leur contenu soit équivalent.
- Les principes à appliquer sont les mêmes que ceux qui ont été mis en place dans le contexte du projet ALICE-T (TEXTERFACE).
- Pour la gestion de la cohérence on disposera des éléments suivants :

##### 1. Champs de gestion :

\* Dans la BDP-Archive :

pour chaque document: CREATION.L: date de création du document en local; UPD.L: date de modification du document en local;

pour chaque document et par version linguistique: TRANSF\_L: date de transfert vers la BDD; CONF\_L: date de confirmation de chargement dans BDD; CREATION.C: date de création du document en central; UPD.C: date de modification du document en central

\* Dans les BDD's :

CREATION avec sous-champs: CREA.A: date de création de la partie analytique; CREA.T: date de création de la partie textuelle du document; et

UPD avec sous-champs: UPD.A: date de modification de la partie analytique; UPD.T: date de modification de la partie textuelle.

Au niveau de la définition de la base, ces champs auront une présentation différente (voir annexe 3).

2. Une table de "présence" en local qui contiendra l'image de chaque base de diffusion et qui permettra de connaître à tout moment si un document existe dans une base. Cette table fera la distinction entre les parties "textuelle" et "analytique". Pour la gestion de la cohérence de cette table on utilisera un processus de consolidation des mises à jour centrales dans le système local.

3. Pour chaque document dans la BDP-Archive sera défini son état dans le processus de production. Ainsi on distinguera les documents en:

- \* documents en production
- \* documents prêts pour transfert
- \* documents transférés pour chargement (dans une version linguistique)
- \* documents déjà chargés dans BDDs (dans une version linguistique).

- Le processus de gestion se déroulera comme suit :

- \* Sélection des documents "candidats" à être transférés vers les BDDs en utilisant comme critères leur état et les dates de gestion.
- \* Après le processus de mise à jour de chaque base de diffusion, extraction du résultat de la mise à jour effectuée et envoi du résultat vers le système local.
- \* Consolidation du résultat de la mise à jour centrale dans l'archive via le NUMERO et le REFER de chaque document ; il sera possible d'établir la relation entre les documents transférés et ceux qui ont été réellement chargés dans les bases.

- L'élimination des documents dans les BDD's sera toujours "contrôlée" à partir du système local et il ne sera pas possible d'éliminer directement les documents dans les BDD's.

#### o. Estimation des volumes

##### - Espace disque

Les estimations suivantes concernent les trois sous-systèmes ALICE-T, ALICE-A et BDDs CELEX.

Sur la base des informations statistiques existantes on s'appuie sur les paramètres suivants :

- Le nombre moyen de documents par an et par version linguistique est de 8.000 (en tenant compte de l'élargissement de la couverture)
- La taille moyenne d'un document est de l'ordre de 10.000 caractères.

- L'espace disque occupé pour l'application CELEX est à l'heure actuelle de l'ordre de 8 GB (toutes les versions linguistiques). Les bases contiennent plus que 130.000 documents (pour les calculs on considère 140.000 documents) et les tailles varient entre 0,7 GB (base espagnole) et 1,1 GB (base néerlandaise).
- L'élargissement de la couverture de CELEX va signifier une augmentation initiale de 10.000 documents avec texte (10.000 caractères par document en moyen) par version linguistique (+2.000 codifiés, +2.000 nouveau Secteur 7, +5.000 saisines, +1.000 rectificatifs) et une élimination de 19.000 documents du Secteur 7 actuel sans texte (500 caractères par document). À ces chiffres il faut rajouter les textes manquants qui seront introduits dans le cadre du sous-projet TEXTERFACE (v. Annexe III de l'étude TEXTERFACE, ref. 2), et la création éventuelle des Secteurs 8 et 10 (2.000 et 10.000 documents de 1.000 caractères respectivement).
- Le nombre maximum de documents en attente de transfert de la partie textuelle vers les BDDs sera de l'ordre de 10.000 (toute version linguistique).

Pour le calcul de volume des bases BDP-Textes et BDP-Archive il a été utilisé une simplification de la formule de calcul fournie par ORACLE (v. ref. 6 et Annexe 7).

= Système local :

\* Système de production de la partie textuelle (BDPT-R, T):

- volume initial : 181 MB
- croissance annuelle : 12,7 MB

\* BDP-Archive :

- Volume initial : 259 MB
- Croissance annuelle : 8,5 MB

\* Gestion :

Pour la gestion de l'application en local une prévision est faite d'un espace de gestion de l'ordre de 300 MB.

Volume total système local :

BDP-Textes :	181
BDP-Archive :	259
Gestion :	400

TOTAL : 840 MB

= Système central :

Pour les estimations sont utilisés les chiffres correspondant aux bases actuelles puisqu'au niveau occupation disque il n'y aura pas de variation importante.

## \* BDD's :

- Occupation actuelle	8.000 MB
- Elargissement de la couverture (10.000 doc. x 10.000 car./doc x 9)	900
- TEXTERFACE	2.518
- Secteurs 8 et 10 (12.000 doc. x 1.000 car./doc x 9)	108

TOTAL : 11.526 MB

Il faut prévoir une croissance annuelle de :  
10.000 doc. x 10.000 car/doc. x 9 = 1.000 MB

## \* Espace de gestion :

L'espace de gestion occupé par CELEX à l'heure actuelle est de l'ordre de 2'5 GB pour les fichiers de travail et de l'ordre de 3 GB pour faire la mise à jour parallèle de 2/3 bases. Donc il faudrait un espace de gestion de l'ordre de 5'5 GB pour faire la gestion des bases en supposant une mise à jour parallèle d'un maximum de 3 bases.

Il faut également prévoir l'augmentation annuelle de cet espace de gestion.

- Impression

On se base sur l'hypothèse qu'après modernisation la gestion de tout le système (local et central) sera effectuée par l'équipe CELEX. Ceci implique un rapatriement des listings du site central vers le site local où s'effectue l'impression.

Les données à imprimer sont en codes 8-bits (ISO 8859/1 pour le latin et 8859/7 pour le grec).

Les besoins de l'équipe CELEX en impression sont évalués comme suit :

	pages/ jour
- validation des textes courantes JO L	2.000
- validation des textes courantes JO C (prévu@)	2.000
- listings mise à jour partie analytique	2.500
- listings mise à jour partie textuelle (800x9)	7.500
- autres listings (versions coordonnées, profils)	5.000
TOTAL	19.000

@ initialement validation du rattrapage



### IV.3 Architecture informatique : configuration HW/SW :

#### a. Systeme local

##### - Hardware

La configuration complète pour la totalité du projet de modernisation a été présentée dans l'étude de faisabilité du premier sous-projet TEXTERFACE (ref. 2).

La configuration demandée est la suivante :

- Une machine avec un système d'exploitation POSIX XPG3 avec :
  - \* espace disque de l'ordre de 1000 MB;
  - \* 1 dérouleur de bandes magnétiques;
  - \* unités d'impression;
  - \* 15 à 20 terminaux

La puissance de la machine doit être telle qu'elle puisse supporter une base relationnelle ORACLE d'un volume de l'ordre de 300 MB et avec des tables de plus de 1 million d'occurrences.

##### - Software

Les éléments software à utiliser sont les suivants :

- Le SGBD relationnel ORACLE version 6 pour la définition/création de la base locale BDP-Archive.
- Le logiciel SQL\*FORMS version 3 pour l'interface d'accès interactif à la BDP-Archive (et aux bases BDPT-A et BDPT-T de ALICE-T).
- Le logiciel MARKIT, version 2.2.
- Le langage C pour le développement de programmes.
- Le logiciel MFTS pour l'échange d'informations avec la machine centrale.

#### b. Systeme central

##### - Hardware

- La machine BULL-DPS90 du Centre de Calcul, avec le système d'exploitation GCOS 8.
- Une fois le système en production, les estimations d'espace disque sont équivalentes à celles faites à l'heure actuelle, plus l'élargissement de la couverture (au total 10 GB). La croissance annuelle prévue est de l'ordre de 1 GB par an.
- La gestion de CELEX dans un espace de gestion exclusif demande un espace disque de l'ordre de 6 GB.
- Il faudra prévoir la mise à disposition d'un espace disque additionnel pendant la durée des travaux de migration.

##### - Software

- Le SGBD documentaire MISTRAL (version 5.3) sera employé pour la mise en oeuvre des bases de diffusion (BDD).

- Programmes de gestion toutes bases pour l'environnement MISTRAL (toolkit).
- Le logiciel de transfert des fichiers MFTS.
- Pour le développement en central on emploiera COBOL et éventuellement C.

c. Observations

- Sur le choix du logiciel documentaire MISTRAL pour les BDD's :

- a) Pour l'implémentation des bases de diffusion il faut faire appel à un système de gestion de base de données documentaire avec des fonctionnalités spécifiques de recherche documentaire sur texte intégral et la capacité de gérer des bases de plus de 150.000 documents accessibles par 3.000 utilisateurs potentiels dont 200 en simultané.

Dans la liste des produits disponibles de la Direction Informatique, on a le choix entre MISTRAL et BASIS. Dans ce cas précis l'option ORACLE n'est pas adaptée et il faudra attendre un certain temps pour avoir des produits pour lesquels l'approche relationnelle, élargie avec des fonctionnalités spécifiques de recherche documentaire, donnera des performances équivalentes aux performances des "vieux" produits documentaires.

- b) Etant donné la situation transitoire vécue actuellement dans l'évolution des bases documentaires, l'option choisie, MISTRAL, nous semble être à l'heure actuelle la plus prudente parce que c'est le progiciel sous lequel fonctionnent actuellement les bases CELEX, avec des fonctionnalités éprouvées et de bonnes performances au niveau de l'interrogation.
- c) Par rapport à l'option BASIS, les arguments qui ont été signalés dans l'EF sur le TEXTERFACE (voir ref. 2 pages 25-26), restent toujours valables.
- d) Il faut signaler que, dans l'architecture proposée par le système CELEX modernisé, le SGBD de diffusion ne constitue plus un facteur de captivité critique étant donné que le système de production est séparé du système de diffusion, et les données sont échangées entre les deux à l'aide d'un format de données standardisé (SGML).

Cependant, d'un point de vue développement, il faut bien isoler les aspects production liés au SGBD de diffusion, et il serait possible de faire une migration vers un autre progiciel à un coût de développement acceptable. Evidemment, dans l'hypothèse d'une migration, il faudrait estimer les coûts de chargement des bases et les coûts de formation des utilisateurs MISTRAL (+/- 3.000) bien que ce coût pourrait être diminué par la mise en place d'interfaces conviviales.

Sur le choix du progiciel ORACLE avec SQL\*FORMS pour le BDP-Archive:

- e) Le progiciel "idéal" pour l'implémentation du sous-système ALICE-A devrait avoir une structure relationnelle avec une interface permettant la saisie et la validation interactive des

textes de longueur indéfinie (entre 250 et 500 caractères) et des champs de nature répétitive (p.e. champ de type descripteur et/ou mot-clé). Dans ce cas les primitives de recherche relationnelle sont suffisantes et il ne faut pas de primitives de recherche documentaire.

- f) Parmi les progiciels disponibles dans la liste des produits, le meilleur choix semble être ORACLE SQL\*FORMS. Il a été la solution retenue lors de l'EF TEXTERFACE (voir ref. 2).
- g) L'interface SQL\*FORMS V. 2 n'est pas actuellement adaptée à la saisie des champs à longueur indéfinie. Ceci est dû, d'une part à la limitation de la longueur d'un champ ORACLE (250 caractères), d'autre part au fait qu'il travaille avec un déroulement horizontal sur une ligne dans un écran de 80 colonnes. Ce problème semble être amélioré dans la version 3 de SQL\*FORMS.
- h) Dans l'implémentation envisagée, et pour pallier les limitations d'ORACLE et de SQL\*FORMS, il faut gérer, dans l'application elle-même :
- \* les champs de longueur supérieure à 250 caractères via une table intermédiaire ;
  - \* la saisie interactive sur plusieurs lignes.
- i) Il faut aussi souligner la captivité des développements faits avec SQL\*FORMS étant donné la philosophie de ce produit. Il est possible que le passage vers un autre progiciel relationnel entraîne la re-écriture de la partie SQL\*FORMS.

Sur le choix de la machine centrale :

- j) La machine retenue pour la mise en oeuvre des bases CELEX est la machine BULL-DPS90 avec le système d'exploitation "proprietary" GCOS8. Dans la stratégie de la Direction Informatique qui prévoit une migration vers des systèmes ouverts (POSIX), la migration de CELEX entraînerait :
- le choix éventuel d'un autre SGBD documentaire (voir points c, d ci-dessous) ;
  - la re-écriture des modules de la partie système centrale : gestion des BDDs, alimentation BDD.

Le tout pourrait signifier un effort de développement de l'ordre de 6 à 9 h/mois.

d. Architecture informatique (voir fig. IV.5)

L'architecture est analogue à celle qui est en train de se mettre en place pour le sous-projet TEXTERFACE.

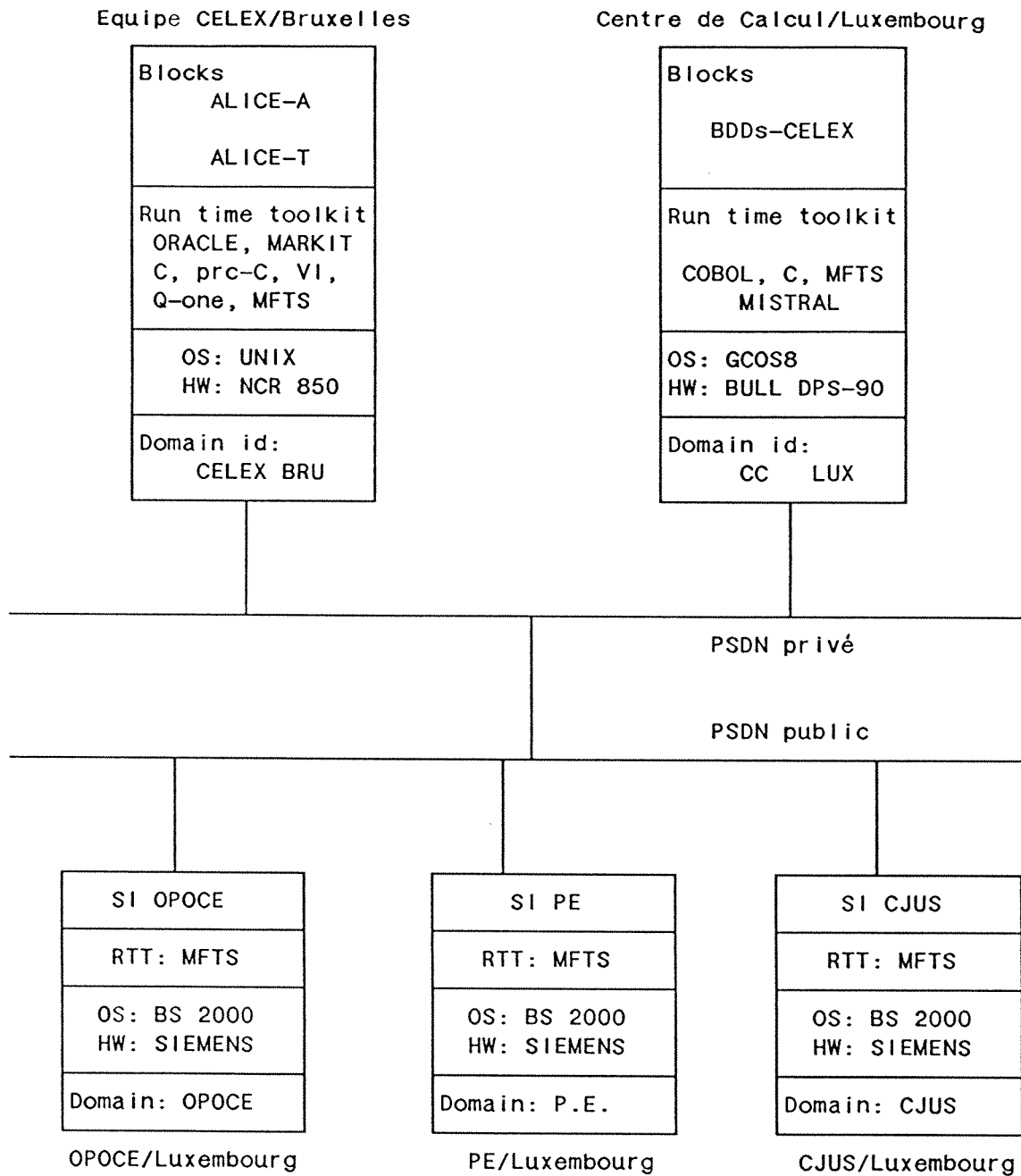


Fig. IV.5 ARCHITECTURE EQUIPEMENT INFORMATIQUE SI CELEX

#### IV.4 Multilinguisme :

Les choix retenus sont en conformité avec ce qui a été déjà établi dans le cadre de l'EF du sous-système ALICE-TEXTERFACE (ref. 2).

##### Format d'entrée

Partie analytique:

PE EBCDIL-SIEMENS

CJ EBCDIL-SIEMENS

Autres institutions

ISO 8859/1 (latin)

ISO 8859/7 (grec)

##### Base de production de la partie analytique (BDP-Archive)

- Champs structurés ISO 8859/1, ISO 8859/7 (tables multilingues)
- Champs textuelles ISO 8859/1 (latin) et ISO 8859/7 (grec)

##### Bases de diffusion (BDDs)

- Bases latines ISO 8859/1
- Base grecque ISO 8859/7

##### Format de transfert BDP-Archive vers BDDs

- Contenu latin ISO 8859/1
- Contenu grec ISO 8859/7

Dans ce cas le latin et le grec pourront coexister dans le même fichier, le contexte pouvant distinguer s'il s'agit de latin ou de grec.

L'annexe VII de l'EF TEXTERFACE (ref. 2) contient en détail toutes les tables citées dans ce chapitre.

#### IV.5 Relation entre les sous-systèmes ALICE-T et ALICE-A :

- Du point de vue échange des données, l'analyse de textes effectuée dans le cadre du système ALICE-T permettra l'extraction d'informations de type analytique qui servira pour initialiser l'analyse documentaire du document dans le système ALICE-A. Ce transfert des données sera fait à l'aide des marqueurs SGML.
- Du point de vue technique, les deux sous-systèmes seront presque totalement intégrés. En fait la BDPT-T et la BDP-Archive seront constituées par un ensemble de tables relationnelles qui feront partie du même dictionnaire.

Certains traitements élémentaires de gestion seront communs et le mécanisme de gestion de la cohérence avec les BDDs sera identique et, conséquemment, le développement de ALICE-A devra être harmonisé avec celui qui a été fait pour ALICE-T.

- Par la consolidation des spécifications des BDDs certains aspects de ALICE-T seront complétés pour gérer les versions codifiées des documents, l'alimentation du Journal Officiel série C et la gestion locale de la documentation en ligne des bases de diffusion.

#### IV.6 Solutions alternatives :

Les solutions alternatives suivantes pourraient être envisagées :

##### a. En ce qui concerne l'architecture de l'application :

1. Installation de la BDP-Archive sur le système central. Cette solution est à écarter parce qu'elle ne suit pas les principes de l'architecture informatique de la Commission et ne réduit en aucun sens le niveau de captivité actuel.
2. Dans le cadre de la solution choisie, on pourrait envisager de générer les parties analytiques correspondant à chaque version linguistique sur le système central. Cette solution n'a pas été retenue car :
  - le volume de transfert de la partie analytique n'est pas trop élevé (1 MB/semaine par mise à jour de toutes les versions linguistiques) par rapport au volume des textes à transférer ;
  - la souplesse et la cohérence de la solution seraient mineures : il faudra avoir la table de correspondance des champs en central, à chaque mise à jour envoyer les tables de traduction par langue ou faire une double saisie tandis qu'on se heurte à une difficulté majeure qui est la gestion de la cohérence
  - une migration suite à des choix éventuels d'autres éléments hardware ou software, serait toujours plus difficile à réaliser.

- ##### b. En ce qui concerne les choix des progiciels, les différentes possibilités sont BASIS au lieu d'ORACLE pour l'Archive et BASIS au lieu de MISTRAL pour les BDD's. Ces alternatives ont été discutées dans le chapitre IV.3.3.

#### IV.7 Evolution technique :

L'évolution technique de la solution décrite est analysée comme conséquence de la disponibilité prévisible à moyen terme des outils hw/sw plus évolués.

##### a. Analyse documentaire en ligne avec visualisation simultanée du texte du document.

Cette fonctionnalité exige un terminal graphique et une interface multi-fenêtre avec la possibilité de visualisation simultanée du texte du document et de l'écran de saisie de l'analyse du document.

#### b. Traitement des tableaux et graphiques.

Les tableaux et graphiques ne sont pas gérés dans le système ALICE étant donné qu'il n'y a pas de définition formelle dans l'entrée des documents (JO) et aussi parce que le SGBD de diffusion ne permet pas leur intégration.

Pour l'introduction du traitement des tableaux dans CELEX, il faudrait :

- l'adoption d'une norme internationale en la matière;
- La disponibilité des terminaux graphiques normalisés et bon marché permettant une visualisation convenable des tableaux et des graphiques.
- Un outil de manipulation et visualisation des tableaux et des graphiques en local, prévu dans le cadre de l'implémentation du step 3 de l'Architecture Informatique de la Commission.
- La gestion des tableaux et des graphiques par le SGBD de diffusion.
- La définition formelle de la structure des tableaux sous FORMEX (ou sous un autre format).

#### c. Traitement morphologique des textes et autres améliorations

Le dispositif envisagé pour le traitement morphologique des textes doit prévoir l'indexation des racines des mots, la vérification des mots contre un dictionnaire, l'activation ou non par l'utilisateur de la recherche en utilisant les racines et les homonymes, utilisation des modèles d'indexation et de recherche avec attribution des poids aux termes du fichier inversé ou aux termes de la recherche, recherche avec analyse du langage naturel et éventuellement des modèles d'aide à la recherche avec "relevance feedback".

#### d. Valorisation des notions hypertexte dynamique

Les liens établis entre les documents de CELEX (actes modifiants, modifications ultérieures, base juridique etc.) pourraient s'enrichir via l'implémentation d'un interface capable d'établir automatiquement des liens supplémentaires entre documents en effectuant des expansions des recherches sur base des descripteurs signalés par l'utilisateur.

### IV.8 Aspects sécurité :

#### a. Système local

- Le système sera accessible par l'équipe de gestion CELEX et par les institutions participant à la production CELEX.
- Il y aura un mécanisme de contrôle d'accès des utilisateurs à plusieurs niveaux (DBM, DBA, analystes/linguistes) tant au système ALICE qu'à la machine elle-même.

b. Système central

- Les BDDs CELEX sont publiques et ne contiennent pas de documents confidentiels. Il y aura seulement des champs de gestion des documents qui ne sont accessibles que par l'équipe de gestion CELEX.
- Le mécanisme de contrôle d'accès pour les utilisateurs externes et internes sera le même que celui qui existe actuellement.



## V. DEVELOPPEMENT ET MISE EN PLACE

### V.1 Stratégie de développement :

#### a. Généralités

Dans la stratégie proposée on a tenu compte des aspects suivants:

- Le processus de migration des bases actuelles demandera un travail important de gestion et il faudra l'accomplir avant l'abandon de l'application ancienne.
- Le sous-projet ALICE-TEXTERFACE déjà développé. Bien que, fonctionnellement, l'interface avec le sous-système ALICE-T soit clairement définie, il faut harmoniser l'implémentation existante avec le nouveau système (configuration de l'application, interface utilisateurs, critères de développement, ...).
- L'environnement de développement et de production est assez compliqué et fait appel à plusieurs technologies et plusieurs progiciels différents.
- La structure modulaire de la solution proposée permettra, d'un point de vue strictement technique, de développer certains modules séparément et de procéder ensuite à leur intégration.

Dans le but de profiter le plus tôt possible des développements achevés, d'accélérer la mise en production, de faciliter le développement et une éventuelle sous-traitance, la stratégie proposée est la suivante :

- Décomposition du projet global (sous-projet 2 de la Modernisation CELEX) en trois hypo-projets : outils de migration, ALICE-A et BDD's-CELEX. On utilise le terme hypo-projet pour ne pas confondre avec les sous-projets du projet global.
- Pour les hypo-projets ALICE-A et BDD's-CELEX, prévoir la livraison et la mise en production échelonnées des développements achevés. Pour gérer cette situation et pour éviter la multiplicité de mises en production, on définira le concept de version du produit qui groupera un ensemble de fonctionnalités prévues d'avance.

#### b. Structure en hypo-projets

On considère les 3 hypo-projets suivants :

##### 1. Outils de migration :

Ce développement demande la connaissance de l'ancienne application (surtout au niveau de l'ancienne ARCHIVE) et il faut prévoir une interaction constante avec l'équipe CELEX au niveau des spécifications pour bien mener la conversion des données.

2. Système local (ALICE-A) :
  - . Il comprend tous les modules liés au fonctionnement du système de production/alimentation de la partie analytique.
  - . L'interface avec les BDDs sera définie à l'aide d'un format de transfert spécifique, défini sur base de SGML.
3. Système central (BDD-CELEX) :
  - . Il comprend tous les modules liés à la gestion des bases de diffusion.

#### c. Gestion des versions

Il est proposé la livraison de 3 versions du système modernisé (ALICE-A et BDDs-CELEX) avec des fonctionnalités croissantes jusqu'à intégrer dans le système toutes les spécifications décrites dans la présente étude de faisabilité.

Chaque version signifie la stabilisation de la précédente plus la mise en oeuvre des nouvelles fonctionnalités prévues.

#### Version 1 :

L'objectif de cette version est de permettre de démarrer les travaux de migration des bases de façon à arriver à des bases nouvelles et anciennes, et des archives nouvelles et anciennes au même niveau des données. Cette version comprendra (point IV) :

- hypo-projet ALICE-A
  - . construction BDP-Archive
  - . saisie interactive
  - . gestion thesaurus
  - . interface ALICE-A/BDD's local minimal
  - . gestion BDP-Archive minimale
- hypo-projet BDDs CELEX
  - . construction BDD's
  - . gestion et alimentation BDD minimale
  - . interface ALICE-A/BDD's central minimal
  - . le système de "document delivery" en liaison avec l'OPOCE

En parallèle avec le développement de cette version il faudra procéder aussi au développement des outils de migration.

Cette version fonctionnera simultanément et sans interférence avec l'application actuelle et elle permettra de tester valablement les nouvelles chaînes.

#### Version 2 :

L'objectif de cette version est de rendre possible l'abandon de l'ancienne application et elle contiendra tous les éléments nécessaires à la production courante. Dans cette version on aura en plus :

- hypo-projet ALICE-A
  - . interface ALICE-A/BDD locale complet
  - . gestion BDP-Archive complète
  - . intégration avec ALICE-TEXTERFACE

- hypo-projet BDDs CELEX
  - . gestion BDD's complète
  - . gestion cohérence BDD/BDP-Archive
  - . diffusion serveurs
  - . interface ALICE-A/BDD centrale complet

### Version 3 :

L'objectif de cette version est de stabiliser la version précédente et surtout d'automatiser les aspects d'administration et de pilotage. Cette version comprendra :

- le module d'administration/pilotage
- l'automatisation de la gestion locale et centrale

#### d. Organisation des travaux

En ce qui concerne l'organisation des travaux de développement, conformément à la politique de la Direction Informatique, il existe les alternatives suivantes :

1. Sous-traitance complète :
  - 1.1 Projet complet
  - 1.2 Organisation en plusieurs hypo-projets
2. Développement interne :
  - 2.1 Avec des ressources internes
  - 2.2 Avec des ressources externes

#### Observations :

- L'alternative 2.1 ne semble guère réalisable à l'heure actuelle vu l'état de la relation projets en cours/ressources humaines disponibles à la Direction Informatique.
- L'alternative 1.1, bien que fort intéressante d'un point de vue formel (un seul responsable contractuel), entraîne la difficulté de trouver un fournisseur capable de disposer et de reproduire tout l'environnement technique (plusieurs progiciels et plusieurs technologies). L'alternative 1.2 permettrait d'éviter ce problème mais il faudrait faire très attention aux problèmes de l'intégration.
- Dans l'hypothèse du choix de l'alternative 1 il faudrait préparer un appel d'offre avec des spécifications détaillées, prévoir un mécanisme de suivi des travaux en cours et tenir compte de la difficulté inhérente à tout système d'information d'arriver à la stabilisation des spécifications utilisateurs.
- L'alternative 2.2 serait intéressante dans l'hypothèse de continuer avec une partie de l'équipe de développement du projet TEXTERFACE.

Par la suite on retiendra l'alternative 1.2.

## V.2 Stratégie de mise en place

CELEX est un système d'information dont le fonctionnement implique de nombreux intervenants, qui est diffusé à un grand nombre d'utilisateurs, et pour lesquels il constitue souvent un outil indispensable de travail. Il faut donc pouvoir mettre en place le nouveau système sans que le service fourni aux utilisateurs soit pénalisé.

La stratégie proposée doit tenir compte du fait qu'il existe des différences importantes entre la structure des anciennes bases et celle des nouvelles et, par conséquent :

- La stratégie proposée (alternative 3) dans la préanalyse (v. ref. 1), qui consiste à faire travailler la nouvelle archive avec les anciennes bases, n'est pas faisable puisque cela nous obligerait à un développement coûteux qui ensuite n'est plus utilisable.
- On ne pourra pas faire de comparaison entre les bases sur le plan de l'exactitude du contenu (du moins en ce qui concerne leur partie analytique), et il faudra plutôt démontrer qu'elles sont équivalentes en ce qui concerne le nombre des documents et, éventuellement, la qualité des stratégies de recherche.
- La garantie de fonctionnement va exiger, pendant une certaine période, la coexistence des nouvelles et anciennes bases.

La stratégie proposée repose sur la conversion progressive de l'ancienne archive vers la nouvelle BDD-Archive et des anciennes bases vers les nouvelles, et comprend les activités suivantes (pour l'enchaînement de ces activités, cf. V.3) :

### a. Création des nouvelles BDD's et chargement initial depuis les anciennes bases.

- Ce chargement comprendrait seulement les champs NUMERO, REFER et les champs textuels.
- Ce processus est simple et automatisable et il n'exige pas de gros effort de développement (de l'ordre de 25h./jour).
- Etant donné que le processus de migration prend un certain temps pendant lequel l'alimentation des anciennes bases CELEX va continuer, la mise à niveau des bases se déroulera de la façon suivante :
  - . prendre l'état actuel d'une base (date t0) et faire le chargement complet de la nouvelle base ;
  - . pour la récupération des chargements faits pendant la durée de chargement antérieur, on a deux possibilités :
    - \* extraire les documents chargés après la date t0 et appliquer le même mécanisme
    - \* utiliser les données stockées dans le système d'alimentation de la partie textuelle ALICE-T

**b. Conversion de l'ancienne archive vers la nouvelle BDP-Archive.**

- Ce processus est automatisable partiellement et, dans certains cas, il faudra une intervention manuelle (à l'aide du module de saisie en ligne).

- Techniquement ce processus comprendra :

- \* le "renumdocage" manuel des documents section 1 ;
- \* le traitement quasi automatique des champs analytiques ;
- \* le traitement des champs de références entre documents (lexique REL). Ici il faut s'attendre à des difficultés de validation.

- Les étapes de cette conversion sont :

- \* Prendre l'état actuel de l'ancienne archive (date t0) et démarrer le processus de conversion.
- \* Quand le processus de conversion de l'ancienne archive est terminé, copier à nouveau l'état actuel de l'archive (date tn) et reconstruire une archive avec tous les mouvements entre la date t0 et la date tn. Appliquer le processus de conversion à cette archive en faisant la distinction entre les documents modifiés et les documents nouveaux.
- \* Terminer le processus antérieur au moment où les deux archives seront au même niveau.

**c. Génération de la partie analytique des nouvelles bases à partir de la nouvelle BDP-Archive.**

- Ce processus peut se faire au fur et à mesure de la construction de la BDP-Archive ou une fois que les deux archives seront au même niveau. Probablement la meilleure solution pour éviter la multiplicité des mises à jour sera de le faire en une seule fois.

**d. Validation du processus de migration des bases.**

A l'aide de l'exécution des procédures automatiques on fera les validations suivantes :

- L'égalité du nombre de documents de la BDP-Archive et de l'archive ancienne (DOC A\_OLD = DOC A\_NEW).
- L'égalité du nombre de documents de la BDP-Archive et des bases nouvelles (DOC A\_NEW = DOC CLX\_NEW).
- L'égalité du nombre de documents des bases anciennes et des bases nouvelles (DOC CLX\_OLD = DOC CLX\_NEW).
- L'équivalence entre les contenus des archives (ancienne et nouvelle).
- L'équivalence entre les contenus des bases (ancienne et nouvelle).

**e. Double saisie et validation des résultats analytiques.**

Il s'agit de procéder à la saisie de la partie analytique dans les deux systèmes et de valider "visuellement" les résultats.

A noter que cette phase pourrait être évitée étant donné que la validation du nouveau système a été réalisée pendant le processus de migration.

Observations :

- La fig. V.1 contient le schéma de la solution proposée. Dans ce schéma on a utilisé un modèle simple du fonctionnement du système CELEX pour montrer l'évolution du processus de migration. La partie gauche correspond au système actuel, la partie droite correspond au nouveau système et la partie centrale contient les actions liées à la migration.
- La solution proposée est une amélioration de la solution 1 de la préanalyse (ref. 1).

V.3 Plan de migration de l'ancien système :

Le chapitre V.2 reprend la description technique de la migration. Ici il s'agit d'estimer les ressources nécessaires et de planifier dans le temps les travaux à effectuer de façon à optimiser l'utilisation des ressources machine et des ressources humaines. Il faut souligner ici l'importance de la migration en tenant compte du fait qu'il s'agit d'une migration de 9 bases, qui correspond à un volume total de l'ordre de 8 GBytes; en plus le système actuel est utilisé en ligne par plus de 2.500 utilisateurs et, par conséquent, il ne faut pas pénaliser la production courante.

L'élaboration d'un plan de migration se fera sur base des aspects suivants :

- Concernant les volumes à traiter et le temps de la migration on dispose les éléments suivantes :
  - \* le nombre de nouveaux documents est de l'ordre de 800 documents par mois et par version linguistique ;
  - \* le volume moyen d'une base CELEX est à l'heure actuelle de l'ordre de 1000 Mb;
  - \* le nombre de documents dans l'archive actuelle est de l'ordre de 140.000 documents ;
  - \* le temps de rechargement d'une base CELEX est de l'ordre de 15 jours elapsed ;
  - \* le temps de régénération d'une base CELEX depuis l'archive est de l'ordre de 15 jours elapsed.
- La garantie de fonctionnement de l'application actuelle et la validation du processus de migration exigera, qu'à un certain moment, les anciennes et les nouvelles bases soient en ligne. L'objectif du plan de migration est de diminuer le temps de disponibilité des deux environnements et d'optimiser l'effort de gestion.

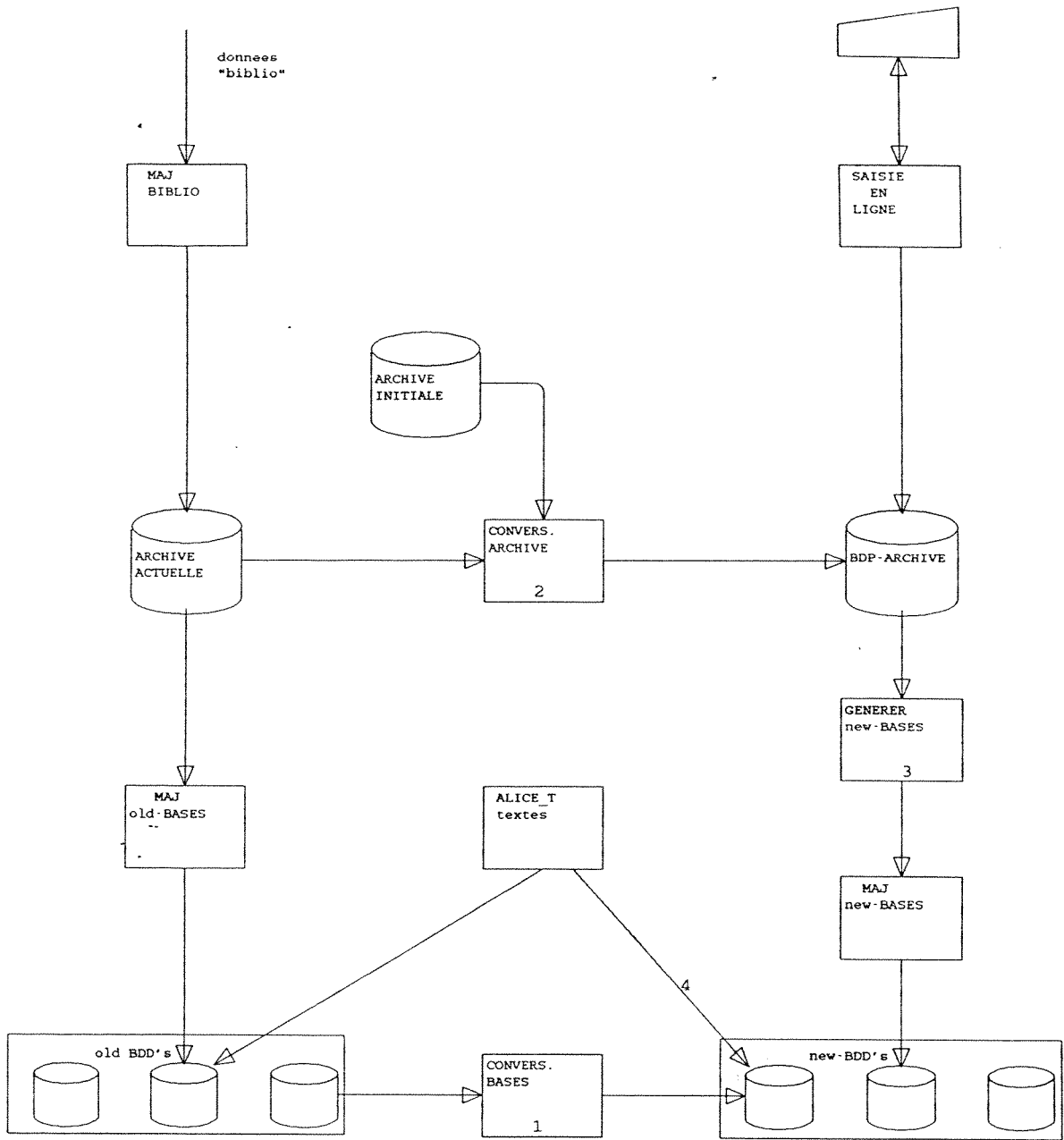


Fig. V.1 PROCESSUS DE MIGRATION

- Les trois "grosses" activités de la migration sont :

- \* la migration de l'archive ancienne (concerne seulement la machine locale) ;
- \* la migration des textes (concerne la machine centrale) ;
- \* la génération de la partie analytique des BDDs sera faite dans la machine locale mais le chargement effectif se réalisera en central.

Par conséquent, la migration de l'archive et la migration des bases ne s'interfèrent pas et utilisent des ressources différentes. La génération de la partie analytique ne peut se faire qu'après la construction complète de la BDP-Archive.

Outre le développement des outils nécessaires pour effectuer la migration, le processus de la migration elle-même sera organisé en deux grandes phases : migration initiale et mise à niveau des données.

a. Phase A : migration initiale :

- Pour démarrer les travaux de migration on prendra l'état (date t<sub>0</sub>) actuel de l'archive et des bases anciennes et on procédera à la conversion vers le nouveau système sans se soucier des mises à jour progressives sur l'ancien système.

- Les actions à entreprendre sont :

A1. Migration ancienne archive.

- \* Ce travail est semi-automatique et il faudra une intervention humaine pour valider les données.
- \* Durée estimée : 7 mois elapsed.
- \* Ressources :
  - gestion informatique : 30 h/j
  - gestion des données : 150 h/j
- \* Comme préalable à cette action il faut la disponibilité de la version 1 (cf. V.1) du nouveau système modernisé.

A2. Migration de la partie textuelle des bases anciennes.

- \* Ce travail est automatique ; la seule restriction est de pouvoir faire la conversion simultanée des 8/9 bases à la fois. On choisira l'hypothèse de faire la migration en 4 blocs (3 blocs de 2 bases et 1 bloc de 3 bases).
- \* Durée estimée : 0,5 m. x 4 = 2 mois elapsed
- \* Ressources :
  - gestion informatique : 20 h/j
  - gestion des données : 18 h/j (2 h/j par base)
- \* Comme préalable à cette action il y a la disponibilité de la version 1 (cf. V.1) du nouveau système modernisé.
- \* Cette action doit être organisée en sorte à ne pas avoir toutes les bases en ligne simultanément et doit démarrer de façon à terminer en même temps que l'action A1.

A3. Génération de la partie analytique des nouvelles bases.

- \* Il s'agit de générer depuis la nouvelle BDP-Archive les parties analytiques à charger dans les nouvelles bases. Par conséquent, l'action A1 doit être terminée.



- \* On adopte également l'hypothèse qu'on procédera au chargement de la partie analytique en 4 blocs.
- \* Durée estimée : 0,5 m. x 4 = 2 mois elapsed
- \* Ressources :
  - gestion informatique : 20 h/j
  - gestion des données : 45 h/j (5 h/j par base)

b. Phase B : Mise à niveau entre le nouveau et l'ancien système.

- Pendant la phase précédente, la production courante a continué. Il s'agit, dans cette phase, de rattraper les textes chargés et les parties analytiques des nouveaux documents créés, et des documents modifiés.

- Etant donné que la durée "elapsed" estimée de l'étape antérieure est de 9 mois, il faudra rattraper les mises à jours de la partie textuelle (4 mois) et analytique (9 mois) faites pendant cette période. Ceci signifie la charge suivante:

- \* partie analytique: 800 documents/mois x 9 mois = 7.200 doc.
- \* textes : 800 documents/mois x 4 mois x 9 (versions)  
x 10.000 caractères/document = 288 MB

- Les actions à faire sont :

B1. Rattrapage des parties analytiques à partir de l'ancienne archive

- \* Durée estimée : 0,5 mois elapsed
- \* Ressources :
  - gestion informatique : 7 h/j
  - gestion des données : 20 jours

B2. Rattrapage des textes

- \* Les textes chargés dans les anciennes depuis que le processus de migration avait commencé
- \* Durée estimée : 0,5 mois elapsed
- \* Ressources :
  - gestion informatique : 7 h/j
  - gestion des données : 9 h/j

B3. Rattrapage de la partie analytique via l'interface BDP-Archive/BDD's

- \* Il s'agit de produire/générer les parties analytiques correspondant au rattrapage de la nouvelle archive.
- \* Durée estimée : 0,5 mois elapsed
- \* Ressources :
  - gestion informatique : 6 h/j
  - gestion des données : 8 h/j

- D'après les estimations faites les actions de rattrapage vont se dérouler dans une période maximum d'un mois. Pendant ce temps la production courante pourrait continuer et répéter le processus de rattrapage jusqu'à atteindre les mêmes niveaux des données. Cependant, la meilleure solution serait de geler la mise à jour des bases CELEX pendant la période de mise à niveau des données de façon à réduire et à simplifier les travaux de gestion.

Plus particulièrement les ressources nécessaires pour procéder à la migration selon la méthode proposée plus haut sont :

Phase	gestion informatique (h/J)	gestion des données (h/J)	temps elapsed (mois)
A1	30	150	7
A2	20	18	2
A3	20	45	2
B1	7	20	0,5
B2	7	9	0,5
B3	6	8	0,5
TOTAL	90 h/J	250 h/J	10,0 @

@ (voir aussi le planning dans le Chapitre X)

La gestion informatique (90 H/J) ainsi que la plus grande partie de la gestion des données (200 H/J) peuvent faire l'objet d'une sous-traitance.

- Ressources en espace disque :

- \* Pendant le processus de migration il faudra l'espace de gestion nécessaire pour traiter deux bases en simultané. Les bases migrées seront sauvegardées jusqu'à ce que la migration d'autres bases soit terminée.
- \* Pour le processus de mise à niveau des données il serait plus facile de pouvoir disposer de tout l'espace nécessaire pour gérer le nouveau système. Cependant il serait possible d'appliquer la même méthode.
- \* Dans l'hypothèse de faire la migration simultanée de 2/3 bases et d'après les volumes de gestion actuels, l'espace disque nécessaire pour la migration est de l'ordre de 3 GBytes.

- Une variante à la méthode proposée serait de disposer et gérer toujours une version linguistique en ligne. Cette variante demanderait d'augmenter l'espace disque et l'effort de gestion, mais avec l'avantage de disposer d'une version pilote que l'on pourrait utiliser pour contrôler la qualité des données.

- Avec cette méthode et le planning proposés, il faudrait avoir les nouvelles et les anciennes bases en ligne pour une période minimum d'un mois. Pendant ce temps il faut disposer de l'espace nécessaire pour avoir en ligne toutes les nouvelles bases, c'est-à-dire qu'il faut 9 GB temporaires.

#### V.4 Estimation détaillée des ressources humaines pour le projet :

L'estimation des ressources humaines nécessaires pour la mise en route du projet, est basée sur l'hypothèse d'une sous-traitance organisée en hypo-projets. On considère les phases suivantes :

- a. Elaboration des spécifications détaillées
- b. Migration
- c. Construction système et acceptance
- d. Running-in
- e. Formation et assistance aux utilisateurs

Pour l'estimation des ressources on considère les hypothèses de travail suivantes :

- On fait la distinction entre les ressources internes (ISI-BDI-I et CELEX-I) qui doivent être disponibles pour garantir le bon déroulement du projet, les ressources externes (ISI-BDI-E et CELEX-E) qui peuvent toujours faire l'objet d'une sous-traitance, et les ressources qui seront demandés au Service de Traduction (CELEX-T).
- Le coût (en hommes/jour) de la participation de l'utilisateur dans les phases de construction et acceptance représentera 15% du coût total de construction. Dans certains cas, et pour des spécifications plus compliquées, ce pourcentage sera de 20%.
- Les ressources internes (ISI-BDI-I) pour la gestion, le suivi et l'acceptance d'un projet développé en sous-traitance seront estimées à 35 % du coût total (en h/j) du projet.

##### a. Elaboration des spécifications détaillées et lancement de l'appel d'offres

- Les spécifications décrites en annexe seront complétées pour procéder à la sous-traitance de la construction du système et également pour consolider les interfaces entre les hypo-projets définis (cf. V).
- Il faudra également inclure les "guidelines" pour le développement, la configuration de l'installation, la liste des documents à fournir, ...
- Il y aura lieu de développer une petite application (10 h/j) pour la saisie des tables de validation et de traduction des champs des bases de diffusion. Ces tables seront annexées à la définition des BDDs.
- Ressources :
  - \* ISI/BDI-I, IDA/IDE : 50 h/j
  - \* CELEX-I : 40 h/j
- Ces ressources comprennent aussi les ressources nécessaires pour le depouillement de l'appel d'offres et la sélection des firmes.

b. Migration

Les ressources pour la migration comprennent deux volets:

- Construction des outils et
- Migration réelle.

Ressources pour la migration

<u>Description</u>	<u>ISI-BDI-E</u> (h/J)	<u>ISI-BDI-I</u> (h/J)	<u>CELEX-I</u> (h/J)	<u>CELEX-E</u> (h/J)
1. Construction				
- Outils archive	70			
- Outils base de diffusion	25			
- Outils de comparaison	15			
Sous-total construction:	110	39	22	
2. Migration réelle	10	3	50	290 (@)
Total migration:	120	42	72	290

(@) Les ressources BDI-CELEX pour la migration réelle sont partagées en gestion informatique (90 h/j) et gestion des données (250 h/j) (v. Table dans Chapitre V.3)

Espace disque nécessaire : 3 GB pendant 12 mois et 9 GB pendant 2 mois.

c. Construction système et acceptance

La liste des actions (structurée par version du système) avec leur estimation en ressources est la suivante :

Ressources pour les actions version 1

<u>Description</u>	<u>ISI-BDI-E</u> (h/J)	<u>ISI-BDI-I</u> (h/J)	<u>CELEX-I</u> (h/J)
1. Hypo-projet ALICE-A :			
- construction base BDP-Archive	30		
- alimentation/saisie interactive	40		
- gestion thésaurus	20		
- gestion BDP-Archive minimale	25		
- interface ALICE-A-BDD local	35		
- documentation	15		
- intégration et tests	15		
- Sous-total 1 : .....	180	63	27
2. Hypo-projet BDD's-CELEX :			
- construction bases	30		
- interface ALICE-A-BDD cent.	15		
- alimentation	15		
- intégration et tests	10		
- documentation	10		
- gestion base minimale	20		
- Sous-total 2 : .....	100	35	15
TOTAL GENERAL :	280	98	42

Ressources pour les actions version 2

<u>Description</u>	<u>ISI-BDI-E</u> (h/J)	<u>ISI-BDI-I</u> (h/J)	<u>CELEX-I</u> (h/J)
1. Hypo-projet ALICE-A :			
- stabilisation version 1	10		
- saisie externe	40		
- interface ALICE-A/BDD's local	25		
- compléter outils gestion DBA	50		
- intégration ALICE-T	20		
- documentation	15		
- intégration et tests	20		
- Sous-total 1 : .....	180	63	27
2. Hypo-projet BDD's-CELEX :			
- stabilisation version 1	15		
- outils de gestion BDD's	40		
- diffusion serveurs	20		
- interface ALICE-A/BDD's	15		
- documentation	15		
- intégration et tests	20		
- Sous-total 2 : .....	125	44	19
<b>TOTAL GENERAL :</b>	<b>305</b>	<b>107</b>	<b>46</b>

Ressources pour les actions version 3

<u>Description</u>	<u>ISI-BDI-E</u> (h/J)	<u>ISI-BDI-I</u> (h/J)	<u>CELEX-I</u> (h/J)
1. Hypo-système ALICE-A :			
- stabilisation version 2	20		
- administration/pilotage	70		
- documentation	15		
- intégration et tests	15		
- Sous-total 1 : .....	120	40	18
2. Hypo-système BDD's-CELEX :			
- stabilisation version 2	15		
- automatisation gestion	60		
- outils de diffusion	60		
- documentation	15		
- intégration et tests	10		
- Sous-total 2 : .....	160	56	24
<b>TOTAL GENERAL :</b>	<b>280</b>	<b>96</b>	<b>42</b>

RECAPITULATION DES RESSOURCES POUR LA CONSTRUCTION DU SYSTÈME  
(par version)

	<u>ISI-BDI-E</u> (h/j)	<u>ISI-BDI-I</u> (h/j)	<u>CELEX-I</u> (h/j)
Version 1			
ALICE-A	180	63	27
BDD's-CELEX	100	35	15
Sous-total 1 : .....	<u>280</u>	<u>98</u>	<u>42</u>
Version 2			
ALICE-A	180	63	27
BDD's-CELEX	125	44	19
Sous-total 2 : .....	<u>305</u>	<u>107</u>	<u>46</u>
Version 3			
ALICE-A	120	40	18
BDD's-CELEX	160	56	24
Sous-total 3 : .....	<u>280</u>	<u>96</u>	<u>42</u>
TOTAL GENERAL : .....	865 h/j	301 h/j	130

La table suivante reprend le même récapitulatif mais par hypo-projet:

RECAPITULATION DES RESSOURCES POUR LA CONSTRUCTION DU SYSTÈME  
(par hypo-projet)

	<u>ISI-BDI-E</u> (h/j)	<u>ISI-BDI-I</u> (h/j)	<u>CELEX-I</u> (h/j)
Hypo-projet ALICE-A :			
Version 1	180	63	27
Version 2	180	63	27
Version 3	120	40	18
TOTAL :	<u>480</u>	<u>166</u>	<u>72</u>
Hypo-projet BDD's-CELEX :			
Version 1	100	35	15
Version 2	125	44	19
Version 3	160	56	24
TOTAL :	<u>385</u>	<u>135</u>	<u>58</u>

Pour le suivi la gestion et la réalisation des tests d'acceptance les temps elapsed prévus sont les suivants :

	<u>temps elapsed (Jours)</u>
Migration	7
Version 1	15
Version 2	30
Version 3	21

Les ressources internes (ISI-BDI-I) pour la gestion, le suivi et l'acceptance sont déjà incluses dans le point précédent.

d. Running-in

On prévoit des RI suffisants pour pouvoir consolider et stabiliser les spécifications utilisateurs.

	<u>temps elapsed (jours)</u>	<u>ISI-BDI-E (h/J)</u>	<u>ISI-BDI-I (h/J)</u>	<u>CELEX-I (h/J)</u>	<u>CELEX-E (h/J)</u>
Migration	7	7	2	14	
Version 1	20	10	9	20	10
Version 2	30	25	9	30	80
Version 3	21	20	7	30	20

e. Formation et assistance aux utilisateurs

- La formation doit viser l'ensemble des utilisateurs du système (utilisateurs finaux, DBA, DBSA/LSA, DBM, linguistes, analystes) et couvrir des aspects spécifiques du système et l'utilisation des divers logiciels de base.
- Pour les estimations qui suivent il a été supposé que la formation à l'utilisation des divers progiciels de base pour le système local a été couverte lors de la mise en production du sous-système TEXTERFACE, et que s'il faut de la formation complémentaire, celle-ci sera donnée dans le cadre du FORUM et avec l'assistance technique de l'unité IUC.
- À un premier niveau, les ressources indiquées sont destinées exclusivement à l'utilisation du nouveau système.
- Le temps de formation et assistance pour l'équipe CELEX en total est donc :

<u>Type de formation</u>	<u>Durée (en jours)</u>
Formation DBA	8
Formation LSA/DBSA	2
Formation analystes/linguistes	5
Assistance complémentaire	35 (sur 3 mois)

TOTAL ..... 50 jours (45 ISI-BDI-E et 5 ISI-BDI-I)

Ce temps sera divisé entre chacun des hypo-projets.

- Le temps de disponibilité de l'équipe CELEX pour cette formation peut être estimé en 120 h/j.

- Concernant la formation des utilisateurs finaux il faut faire les observations suivantes:

\* La nécessité de former les utilisateurs au nouveau système découle du besoin de promouvoir le nouveau système afin de garantir son acceptation auprès de ses utilisateurs. En fait, comme il a été montré dans le cadre de l'opération "prototype", toutes les nouveautés introduites dans le système faisaient l'objet d'une demande pressante de ces mêmes utilisateurs, qui trouveront dans le nouveau système des réponses à de "vieux problèmes".

\* Les actions plus spécifiques à prévoir sont:

- disponibilité auprès des utilisateurs et en ligne des manuels pour le nouveau système avant son ouverture;  
ressources:  $200 \div 5 = 40$  hommes/jours par version linguistique ou 360 h/j en total
- disponibilité d'un didacticiel sur PC pour apprentissage du système avant son ouverture; (actuellement il fait l'objet d'une sous-traitance indépendante);  
ressources: pour la mise au point du didacticiel sur base de la nouvelle structure des BDD et la traduction dans toutes les langues: 360 h/j
- prévision des cours d'une demi-journée pour la mise à niveau des utilisateurs qui le souhaitent; dans les 2 salles du FORUM on pourrait alors donner 4 cours pour 15 personnes par jour c.à.d on pourrait former 300 personnes par semaine pour toutes les versions linguistiques;  
ressources (en supposant que la moitié seulement des utilisateurs actuels aura vraiment besoin d'un cours de mise à niveau - les autres pourront utiliser le manuel ou le didacticiel) 30 h/j.



## VI. ASPECTS ORGANISATIONNELS

D'un point de vue organisationnel, il faut d'abord signaler que le nouveau système est fondé sur l'idée de la décentralisation de la production de la partie analytique et de la partie textuelle de CELEX.

Il est évident qu'avec la mise en place du projet de modernisation, la production CELEX sera réalisée en site local et la responsabilité du DBM sera élargie suite à l'incorporation de la gestion informatique dans l'équipe CELEX. Egalement, et dans le but de tirer un meilleur profit du nouveau système, il est prévu de réorganiser l'équipe de gestion actuelle par une nouvelle distribution des tâches. Dans cette optique il faudra distinguer les fonctions de la gestion des données de CELEX, de la formation et du contrôle de la qualité linguistique des données.

Ainsi un plus grand effort de management se traduira en une augmentation importante de la qualité et de l'actualité des données.

L'organisation de l'équipe DBM prévu après la modernisation est repris dans l'Annexe 8.

Plus particulièrement les différentes fonctions dans le cadre de l'équipe CELEX peuvent être décrites comme suit:

### a. Gestion informatique

Les estimations de la charge de travail, pour assurer la gestion informatique du système CELEX complet, sont les suivantes :

ALICE-T	0,7 - 1 H/A
ALICE-A	0,3 - 0,5 H/A
BDD's-CELEX	0,6 - 0,8 H/A

Les activités de gestion et les responsabilités correspondent à celles décrites dans le Guide Informatique (doc. Management des bases de données GI 7.2/3-F) pour le profil DBA. Cependant il faut y inclure

- \* la responsabilité de la gestion de logiciel applicatif fourni (celui-ci intégré dans plusieurs logiciels de base) pour l'exploitation de l'application
- \* l'assistance spécifique dans le fonctionnement du système d'information CELEX aux utilisateurs concernés, c.à.d. DBM, analystes et linguistes.

Par conséquent, dans le contexte de la modernisation CELEX, il faut prévoir 2 postes DBA pour la gestion informatique du système complet (partie locale et centrale).

A l'heure actuelle un des deux postes sollicités est couvert, il faut prévoir le recrutement d'une autre personne.

Etant donné la complexité de la structure logique de l'information contenue dans CELEX et celle qui dérive de l'implémentation faisant appel à l'intégration de plusieurs logiciels, il est absolument nécessaire d'exiger un bon niveau de qualification technique pour les DBAs pour qu'ils puissent assurer l'assistance aux utilisateurs du système CELEX (DBM, analystes et linguistes).

Il convient aussi de signaler que la taille du système exige actuellement et exigera encore plus dans le futur jusqu'à la stabilisation du système et le "rodage des utilisateurs", une assistance continue dans le processus de production. Cette assistance sera assurée dans un premier temps par l'équipe de gestion de la production courante.

- Les fonctions LSA ainsi que la maintenance et l'assistance technique sont assurées par l'unité IUC.

b. Coordination documentaire et dispatching

Les tâches de réception dispatching des documents papier et outils d'analyse ainsi que la coordination et validation centrale des documents et des fichiers de travail (num. doc, BN en ligne, fichiers textuels), seront assurées par l'équipe actuelle de gestion 2 C et 1 D (personnel en place) avec une charge de travail de l'ordre de 40 %.

c. Traitement des différentes versions linguistiques

A part la vérification du splittage, le contrôle de la qualité des textes ainsi que les modifications des textes codifiés peuvent être confiés au Service de la traduction comme sous-traitance interne.

VII. ANALYSE COÛTS/BÉNÉFICESVI.1 Coûts :a. Coûts du projet

## - Construction et mise en production

Activité	ISI-BDI-E (h/J)	ISI-BDI-I (h/J)	CELEX-I (h/J)	CELEX-E (h/J)	CELEX-T (h/J)
a. Préparation spécifications et appel d'offre		50	40		
b. <u>Migration</u>					
- Préparation	117	41	36	-	
- Réalisation	-	-	50	290	-
- Formation/ assistance	10	3	-	-	-
Sous-total	127	44	86	290	0
c. <u>Version 1</u>					
- Construction	280	98	42	-	-
- Running-in	10	9	20	10	-
- Formation/ assistance	15	2	40	-	-
Sous-total	305	109	102	10	0
d. <u>Version 2</u>					
- Construction	305	107	46	-	-
- Running-in	25	9	30	80	-
- Formation/ assistance	20	2	60	-	-
Sous-total	350	118	136	80	0
e. <u>Version 3</u>					
- Construction	280	96	42	-	-
- Running-in	20	7	30	20	-
- Formation/ assistance	10	1	20	-	-
Sous-total	310	104	92	20	0
f. <u>Diffusion nouveau système</u>					
- Formation des utilisateurs finaux					30
- Adaptation manuels					360
- Adaptation didacticiels					360
Sous-total	0	0	0	0	750
TOTAL GENERAL	1092	425	456	400	750

soit en H/A **15,7** = 5,5 + 2,1 + 2,3 + 2 + 3,8

Le coût total du projet qui s'élève à 15,7 H/A est ventilé comme suit:

5,5 + 2 = 7,5 H/A en sous-traitance externe (750.000 ECU)

3,8 H/A en sous-traitance Traduction

2,1+2,3 = 4,4 H/A en ressources internes IX/I (voir par la suite l'évolution de l'effectif)

#### - Infrastructure hw/sw

##### \* centrale :

il faut prévoir une augmentation de 2 GB pour la production courante, et la disponibilité de 3 GB pendant 12 mois et 9 GB pendant 2 mois pour la migration. D'après le coût de MB pour la machine BULL (7000 ECU/GB) où le coût total est : 17000 ECU/an plus environ 36000 ECU pour la migration.

##### \* locale :

Dans le contexte du sous-projet TEXTERFACE, la machine suivante a été installée :

- une machine NCR (appelée minos) avec le système UNIX 3.0;
- 3 disques de 380 MB chacun
- 16 terminaux
- 3 imprimantes

Cette configuration doit être complétée par une imprimante multilingue rapide.

#### b. Coûts opérationnels

Après la modernisation on prévoit les coûts suivants:

- Support/maintenance du LSU UNIX minos (IUC);
- Support/maintenance du CSU GCOS 8 BULL (CC);
- Espace disque: total 31,5 GB; prévision d'augmentation annuelle 1 GB;
- Support/maintenance application complète (ALICE-T inclus): 1 homme/année par an. Etant donné la dynamique évolutive de CELEX, il nous semble difficile de diminuer cette prévision. Evidemment cette maintenance pourrait être sous-traitée (coût 100.000 ECU par an).
- Communications: Le fonctionnement du système en régime "croisière" signifie un transfert situé entre 3-6 MB/j.

- Management de l'application/Gestion des données:

Différentes tâches auront changé de nature permettant par la même une restructuration de l'équipe DBM et l'intervention systématique de l'équipe développement. Après modernisation les tâches de gestion du système modernisé demanderont les ressources suivantes:

Equipe DBM 17 H/A et 252.000 ECU

Comme il a été déjà mentionné, cette gestion serait assurée de façon indépendante de toute intervention d'un service ou agent autres. Le système pourrait être étendu avec un minimum d'effort et être converti en central avec peu de ressources également. Il faut aussi mentionner une sous-traitance interne pour le suivi de la qualité linguistique, la production des versions coordonnées, la production des manuels, la diffusion et la formation.

c. Mise en oeuvre et évolution: effectif et espace disque

EFFECTIF

Aujourd'hui:

Ressources internes

Ressources externes

- Equipe de DBM :

3	x A	180.000 ECU (analyse)
5,75	x LA	112.723 ECU (encodage)
8	x B/C	
1	x D	
-----		-----
17,75	H/A	292.723 ECU

- Services associés :

. Diffusion	6 H/A	
. Développement	1 x A	70.000 ECU
. Gestion		2 H/A
. Publication		600.000 ECU

Pour pouvoir avoir une image réelle de l'évolution de l'effectif, il faut se rappeler du fait que l'équipe DBM CELEX tourne actuellement avec un déficit de 4 H/A (4 LAs et 1 B) en personnel.

Pendant la modernisation:

Le présent projet de modernisation exige une augmentation temporaire des effectifs DBM et développement.

Ressources internes                      Ressources externes

- Equipe DBM

Engagement	1 x A management modernisation 2 x B gestion locale	
Contrat externe pour:		
1 H/A gestion des données migration		100.000 ECU
1 H/A gestion informatique migration et assistance informatique pour le running-in		100.000 ECU
Traduction	3,75 H/A	

- Développement

Sous-traitance externe pour 5,5 H/A		550.000 ECU
-------------------------------------	--	-------------

Il est supposé que le reste, à savoir:

Equipe ISI-BDI 2,1 H/A

sera couvert par des ressources existantes.

Indépendamment de la modernisation:

Engagement	3 x A (management production, juriste-analyste, juriste-repertoire) + regularisation 1 LA + conversion 1 B en A 1 x A (management UFI) 9 x A/LA/B (qualité des données) 9 x A/LA/B (codification-documentation) 1 x B (Serveurs) + regularisation 1 C 2 x C (secretariat)
------------	--

plus l'équivalent de 9 A/LA/B pour la formation

ESPACE DISQUE

Aujourd'hui:	Bases	7,1 GB	
	ARCHIVE	0,9 GB	
	Gestion 2/3	5,5 GB	TOTAL: 13,5 GB
Projet:	Augmentation couverture	11,5 GB	
	Migration (pour 12 mois)	3 GB	
	Migration (pour 2 mois)	9 GB	
Après:	Bases	19,5 GB	
	Gestion 9	12 GB	TOTAL: 31,5 GB

AUGMENTATION NETTE : 18 GB  
PENDANT PROJET PREVISION DE 6,5 GB SUPPLEMENTAIRES.  
ENSUITE PREVISION DE 1 GB PAR AN

Pour une analyse de l'organisation de l'équipe CELEX après la  
Modrnisation voir l'Annexe 8.

## VI.2 Bénéfices :

Les bénéfices de la modernisation de CELEX seront très importants à tous les niveaux :

- Les ressources consommées après la modernisation par GByte seront:

$$52 \text{ h/année} \div 31,5 \text{ GB} = 1,65 \text{ h/a par GB}$$

ce qui est équivalent à ce qui est consommé aujourd'hui mais avec une meilleure service aux utilisateurs;

- Le temps nécessaire pour se former à CELEX se réduira à 1 jour de formation au lieu de 2. Ceci aura, hors l'économie de 50 % en ressources formateur, des répercussions immédiates sur l'utilisateur final. Il a été estimé (v. réf. 8) qu'un cabinet juridique moyen situé à Francfort dépenserait de 30% à 100% moins en frais de formation (30% moins pour une formation à Bruxelles, 100% moins pour un cour de perfectionnement qu ne serait plus nécessaire et 65% moins pour une formation sur place).
- Les délais de m.à.j. de la partie analytique et du texte se réduiront à 1 jour en vitesse de croisière; en comparant avec les délais actuels ceci représente une accélération de 86 à 93 %
- La qualité des données sera améliorée à 100% vu le fait que les corrections seront possibles dans les textes contrairement à aujourd'hui;
- Meilleure qualité de l'information disponible dans la base de diffusion due aux contrôles rendus possibles par les outils qui seront mis en place ; gestion des textes plus souple.
- Augmentation de la productivité du personnel par la mise à sa disposition d'outils plus performants ;
- Moins de coûts de maintenance puisqu'une partie de la maintenance sera assurée en local ;
- Possibilité d'intégration dans l'environnement bureautique de la Commission et des autres institutions avec des économies énormes d'encodage des textes modifiés (propositions adoptées puis modifiées). Par exemple , en considérant un ré-encodage de quelques 750 propositions de la Commission au Conseil d'une longueur moyenne de 10.000 caractères dans les 9 langues (67,5 Mbytes) par an, on obtient une économie de frappe de 8,45 hommes/année (20 pages de 2.000 caractères par jour de frappe d'une secrétaire) ;
- Diminution des délais de disponibilité des versions codifiées des actes de la législation (de 2 ans à quelques semaines) ;

- Les vrais bénéficiaires pour les utilisateurs en matière de recherche plus souple et plus rapide, c'est-à-dire moins coûteuse, seront considérables. Il a été estimé (v. ref. 8) que pour le système modernisé dans le cas d'un besoin de reconstitution à leur état actuel d'une série de textes législatifs (p.ex. documents de 1983 sur le vin en vigueur) un cabinet juridique économiserait 70% de coûts (temps de connexion réduit et moins des coûts de prestation de service aux clients finaux). Même si nous prenons l'hypothèse que grâce à la nouvelle structure de la base seule la visualisation de l'information pertinente sera accélérée de 2 minutes sur une recherche de 10 minutes (p.ex. grâce au splittage des champs des textes de la législation) le bénéfice en temps de connexion est de l'ordre de 20%
- En plus l'extension de la couverture aux actes de gestion courante, en combinaison avec la mise à jour rapide de la base, vont certainement conduire à:
  - \* un élargissement important de la clientèle et du temps d'utilisation;
  - \* une diminution des demandes des JOs sur support papier comme les utilisateurs de CELEX auront la possibilité d'être informés sélectivement sur les documents qui les intéressent avant de passer des commandes on- ou off-line;
- L'implémentation du modèle des systèmes ouverts et du SGML dans le cadre du système CELEX va diminuer la captivité de l'application des progiciels et technologies utilisés actuellement; cet argument seul peut justifier en fait l'investissement demandé pour la mise en oeuvre du projet.



### VIII. RESPONSABILITÉS ET SERVICES PARTICIPANTS

Project owner : Le Directeur de l'Informatique

Project leader : Le DBM CELEX

System supplier : La section BDI de l'unité ISI

Services participants au projet :

Développement : Section BDI de l'unité ISI  
Système local : Unité IUC  
Système central: Centre de Calcul et unité Support Logiciels  
Running-in : DBM CELEX avec son équipe  
Formation : Section Forum, IUC

System owner : Les institutions et le Groupe informatique  
juridique du Conseil

Service/data base manager : le DBM CELEX

Service administrator: Intégré dans équipe CELEX

Maintenance : Section BDI de l'unité ISI  
le DBM CELEX avec son équipe

Institutions concernées dans l'alimentation :

- Cour de Justice
- Parlement Européen
- OPOCE (Office des Publications)
- CES (Comité Economique et Social)
- Conseil des Ministres des Communautés Européennes
- Le Secretariat Général de la Commission

## IX. FACTEURS DE RISQUE

### Risque 1 : abandon du projet.

Conséquences :

- Continuation de la situation actuelle de gestion complexe avec la dégradation lente mais progressive de la qualité et de l'actualité.
- L'inadaptation croissante du système aux structures de l'architecture informatique de la Commission.
- L'emploi injustifié de ressources pour la maintenance et l'exploitation d'un système dépassé.

### Risque 2 : retard dans la mise en route du projet par manque de ressources.

Conséquences :

- Une perte de crédibilité de la Commission vis-à-vis du Groupe Informatique Juridique du Conseil et des Institutions et du Parlement Européen qui pourra avoir des implications politiques et contractuelles (envers nos serveurs) graves.
- Une perception négative de la part des milieux professionnels concernés et du public européen en général.
- L'aggravation de la manque de transparence dans le processus législatif communautaire qui aura des implications néfastes quant au fonctionnement démocratique des Institutions de la Communauté après 1993.

### Risque 3 : mauvaise performance de la configuration locale/centrale installée.

Plusieurs problèmes sont à signaler ici :

- l'opération des machines UNIX utilisées à la Commission ne peut pas être considérée comme étant toujours fiable; le premier sous-projet de la Modernisation de CELEX (TEXTERFACE) a connu des énormes problèmes (mauvaise collaboration ORACLE avec MFTS, mauvaises paramétrages de la machine par l'équipe LSA, introduction d'applications non prévues initialement, non attribution d'espace disque etc.).
- la mise en oeuvre d'un système aussi compliqué que le CELEX modernisé, demande la synergie de plusieurs services différents qui doivent disposer des ressources nécessaires à temps; ceci n'est pas évident surtout dans la Direction Informatique qui fonctionne avec pénurie de ressources.

Conséquences :

- Mauvais temps de réponse, incohérences dans les fichiers (quand on change les permissions), perte des fichiers (en cas de saturation de l'espace alloué), etc.

- Perception décevante du projet de modernisation par les utilisateurs; frustration au niveau de l'équipe de DBM.
- Implications considérables sur d'autres projets de la Commission pour lesquels la modernisation CELEX serve de modèle.

Risque 4 : Disponibilité et support convenables du matériel et des progiciels utilisés

La machine NCR 850 ne sera plus supportée par NCR (pas de migration en UNIX V.4).

Progiciels critiques pour le projet:

- MFTS (sous UNIX, BS 2000 et GCOS 8)
- MARK-IT (sous UNIX et GCOS 8)
- ORACLE (sous UNIX)
- Q-Office (sous UNIX)
- MISTRAL (sous GCOS 8)
- xpad (sous UNIX)

La non disponibilité de ces progiciels, leur support partielle, ou leur mauvaise "collaboration" en parallèle, aura les mêmes conséquences que les conséquences mentionnées sous le point 3 supra.

Risque 5 : manque de rigueur dans la gestion et le suivi du projet.

Il faut surtout mentionner ici la non disponibilité des ressources nécessaires pour l'avancement du projet et/ou l'exploitation du système opérationnel à temps:

Conséquences :

- Impossibilité de gérer la complexité du projet, les objectifs compliqués et de l'environnement technique.
- Retard dans le planning prévu.
- Gaspillage des ressources pour construire un système qui ne sera pas utilisé par la suite à son potentiel complet.

Risque 6: Problèmes avec la stabilité du format FORMEX

Conséquences:

- retard dans les m.à.j. à cause des délais nécessaires pour effectuer des validations et corrections supplémentaires;

## X. PLAN DU PROJET

- Le schéma X.1 reprend le planning du projet et l'enchaînement des activités suivant la stratégie définie (cf. fig. V.1). La fig. IX.2 contient le plan RMS.
- Pour simplifier la présentation on a regroupé certaines activités du développement décrit.
- Pour la réalisation de ce planning on a fait les hypothèses de travail suivantes :
  - . d'après le plan de migration (cf. V.4) il faut 10 mois elapsed pour compléter la migration ;
  - . pour les développements des versions proposées (cf V.1) on va supposer les durées suivantes (avec l'ajustement de ressources nécessaires) et supposer que les développements de ALICE-A et BDD's peuvent être réalisés en parallèle :

Outils de migration :	4 mois
Version 1 :	6 mois
Version 2 :	6 mois
Version 3 :	6 mois
- Plus particulièrement le planning suivant contient la séquence des actions. On peut observer que le processus de migration pourrait se dérouler dans un temps elapsed de 6 mois.

14 DEC 90

ACTIVITY DETAIL  
Modernisation CELEX

PAGE 1

## Original Dates and Usage

-----Name-----	Start	End	Duration Bus days	-----Resource assignments----- Name	Usage
EF Approbation EF	1-01-91	1-01-91	1	Milestone	1.0
APPEL D'OFFRE					
AOPE Préparation spécifications	1-01-91	25-02-91	40	ISI-BDI interne BD-CELEX interne	45.0 35.0
AOAM Appel au marché	26-02-91	18-03-91	15	attente	
AORC Réponse à la consultation	19-03-91	15-04-91	20	attente	
AODS Dépouillement et sélection	16-04-91	6-05-91	15	ISI-BDI interne BD-CELEX interne	5.0 5.0
AODD Délai démarrage de contrat	7-05-91	27-05-91	15	attente	
PREPARATION MIGRATION					
MICO Construction outils	28-05-91	16-09-91	80	ISI-BDI externe ISI-BDI interne BD-CELEX interne	110.0 34.0 19.0
MIAC Acceptance	17-09-91	23-09-91	5	ISI-BDI interne BD-CELEX interne	5.0 3.0
MIRI Running-in	24-09-91	7-10-91	10	ISI-BDI interne ISI-BDI externe BD-CELEX interne	2.0 7.0 14.0
MIMEP Mise en production	8-10-91	8-10-91	1	Milestone	1.0
MIPR Assistance migration	9-10-91	29-10-91	15	ISI-BDI interne ISI-BDI externe	3.0 10.0
MIGRATION INITIALE					
MIG11 Migration ancienne archive	9-10-91	21-04-92	140	BD-CELEX externe BD-CELEX interne	150.0 30.0
MIG12 Migration bases	26-02-92	21-04-92	40	BD-CELEX externe BD-CELEX interne	34.0 4.0
MIG13 Génération biblio	22-04-92	16-06-92	40	BD-CELEX externe BD-CELEX interne	56.0 9.0
MIGRATION FINAL ( rattrapage)					
MIG21 Rattrapage archive	17-06-92	30-06-92	10	BD-CELEX externe BD-CELEX interne	24.0 3.0
MIG22 Rattrapage textes	17-06-92	30-06-92	10	BD-CELEX externe BD-CELEX interne	14.0 2.0
MIG23 Rattrapage biblio	1-07-92	14-07-92	10	BD-CELEX externe BD-CELEX interne	12.0 2.0

14 DEC 90

ACTIVITY DETAIL  
Modernisation CELEX

PAGE 2

## Original Dates and Usage

-----Name-----	Start	End	Duration Bus days	-----Resource assignments----- Name	Usage
VERSION 1					
VICOA Construction ALICE-A V1	28-05-91	11-11-91	120	ISI-BDI externe ISI-BDI interne BD-CELEX interne	180.0 57.0 27.0
VICOB Construction BDD's V1	28-05-91	11-11-91	120	ISI-BDI externe ISI-BDI interne BD-CELEX interne	100.0 31.0 15.0
VIAC Acceptance V1	12-11-91	25-11-91	10	ISI-BDI interne	10.0
V1RI Running-in V1	26-11-91	16-12-91	15	ISI-BDI externe ISI-BDI interne BD-CELEX interne BD-CELEX externe	10.0 9.0 20.0 10.0
VIFA Formation et assistance V1	26-11-91	30-12-91	25	ISI-BDI externe ISI-BDI interne BD-CELEX interne	15.0 2.0 40.0
V1HEP Mise en production V1	17-12-91	17-12-91	1	Milestone	1.0
VERSION 2					
V2COA Construction ALICE-A V2	12-11-91	27-04-92	120	ISI-BDI externe ISI-BDI interne BD-CELEX interne	180.0 57.0 27.0
V2COB Construction BDD's V2	12-11-91	27-04-92	120	ISI-BDI externe ISI-BDI interne BD-CELEX interne	125.0 39.0 19.0
V2AC Acceptance V2	28-04-92	18-05-92	15	ISI-BDI interne	11.0
V2RI Running-in V2	19-05-92	13-07-92	40	ISI-BDI externe ISI-BDI interne BD-CELEX interne BD-CELEX externe	25.0 9.0 30.0 80.0
V2FA Formation et assistance V2	19-05-92	3-08-92	55	ISI-BDI externe ISI-BDI interne BD-CELEX interne	20.0 2.0 60.0
V2HEP Mise en production V2	15-07-92	15-07-92	1	Milestone	1.0
VERSION 3					
V3COA Construction ALICE-A V3	16-07-92	30-12-92	120	ISI-BDI externe ISI-BDI interne BD-CELEX interne	120.0 36.0 18.0
V3COB Construction BDD's V3	16-07-92	30-12-92	120	ISI-BDI externe ISI-BDI interne BD-CELEX interne	160.0 50.0 24.0
V3AC Acceptance V3	31-12-92	20-01-93	15	ISI-BDI interne	10.0
V3RI Running-in V3	21-01-93	3-03-93	30	ISI-BDI externe ISI-BDI interne BD-CELEX interne	20.0 7.0 30.0

14 DEC 90

ACTIVITY DETAIL  
Modernisation CELEX

PAGE 3

Original Dates and Usage

-----Name-----	Start	End	Duration Bus days	-----Resource assignments----- Name	Usage
				BD-CELEX externe	20.0
V3FA Formation et assistance V3	21-01-93	24-03-93	45	ISI-BDI externe	10.0
				ISI-BDI interne	1.0
				BD-CELEX interne	20.0
Mise en production V3	4-03-93	4-03-93	1	Milestone	1.0
DIFFUSION NOUVEAU SYSTEME					
PR1 Arrêt ancien système	16-07-92	16-07-92	1	Milestone	1.0
PR2 Système en fonctionnement	17-07-92	8-10-92	60	BD-CELEX interne	30.0
PRMD Manuel et didacticiel	28-05-91	27-04-92	240	BD-CELEX Traduction	720.0
Total project	1-01-91	24-03-93	582	Total days	3129.0

14 DEC 90

ACTIVITY DETAIL  
Modernisation CELEX

PAGE 4

Original Dates and Usage

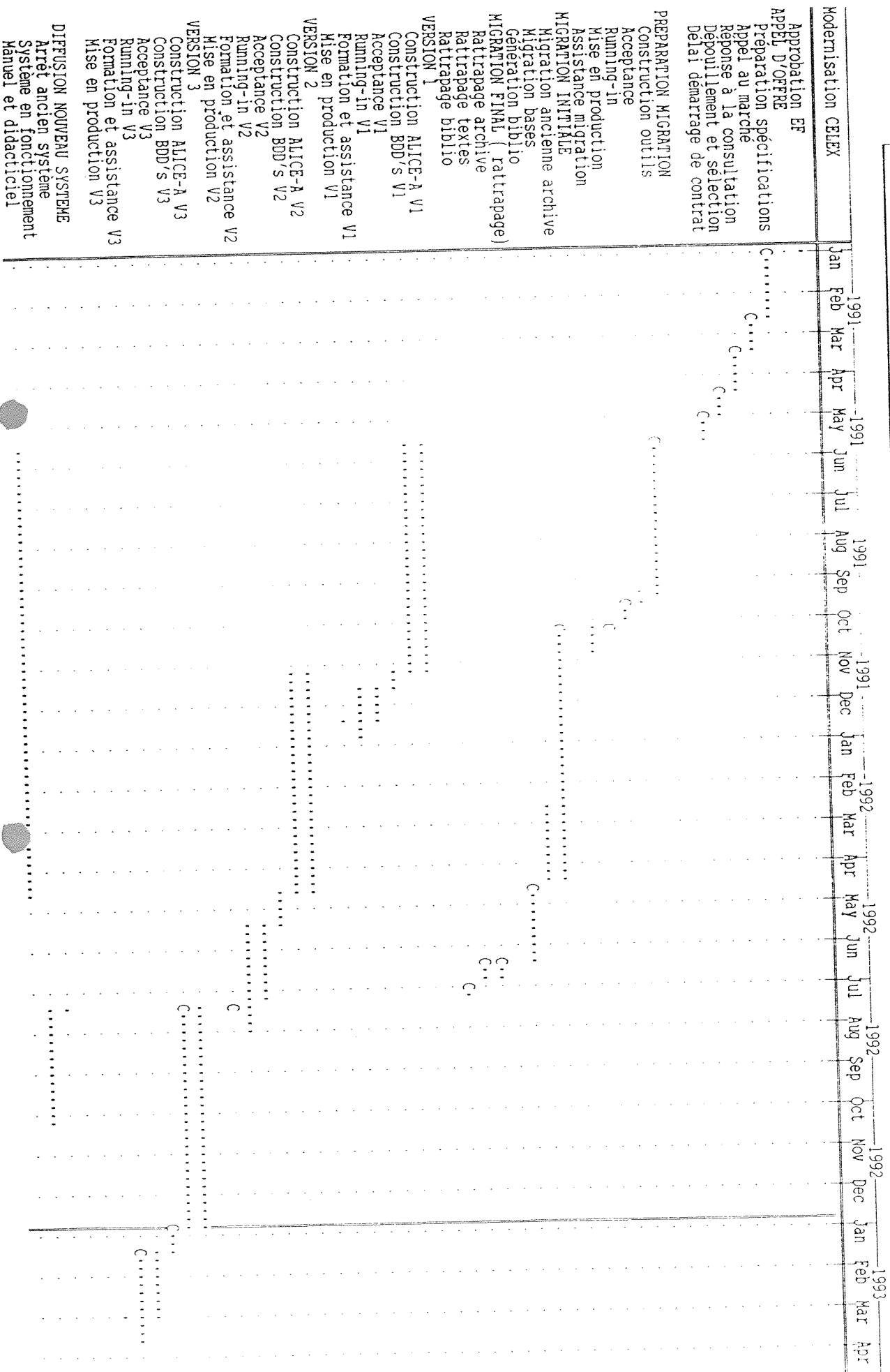
-----Name-----	Start	End	Duration Bus days	-----Resource assignments----- Name	Usage
Resource summary					
Name	Units	Amount			
ISI-BDI externe	days		28-05-91 24-03-93	IE	1092.0
ISI-BDI interne	days		1-01-91 24-03-93	II	425.0
3D-CELEX interne	days		1-01-91 24-03-93	BI	486.0
3D-CELEX externe	days		9-10-91 3-03-93	BE	400.0
3D-CELEX Traduction	days		28-05-91 27-04-92	BT	720.0
Total project			1-01-91 24-03-93 582	Total days	3123.0



14 DEC 90

GANTT CHART  
Modernisation CELEX

PAGE 1



14 DEC 90

GANTT CHART

Modernisation CELEX

PAGE 2

Resource summary Utilization	1991		1991		1991		1991		1991		1991		1992		1992		1992		1992		1992		1993								
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr			
ISI-BDI externe					15.2	76.0	89.7	80.5	66.1	69.4	57.7	77.0	58.9	50.0	55.0	47.5	9.9	22.1	41.5	50.7	51.3	50.6	48.3	50.6	7.0	17.6	5.4				
ISI-BDI interne		26.3	18.7		3.8	6.4	25.6	27.6	25.2	17.8	29.6	27.4	18.4	16.0	16.5	15.7	12.2	5.6	11.4	15.5	15.4	15.4	14.7	16.1	12.1	4.6	0.6				
BD-CELEX interne		20.7	14.3		3.8	4.4	16.0	13.3	11.0	18.5	19.0	25.6	62.8	13.8	11.4	13.2	12.3	21.6	48.1	43.5	19.9	19.8	10.1	6.3	6.6	4.9	13.3	1.3			
BD-CELEX externe										18.7	25.9	31.4	25.3	24.4	41.3	36.8	47.4	98.8	30.0												
Milestone			1.0							1.0																					
attente																															
BD-CELEX Production					12.0	60.0	69.0	66.0	63.0	69.0	63.0	66.0	69.0	60.0	60.0	66.0	57.0														
Total days		48.0	33.0		7.6	38.0	177.6	196.6	183.9	172.8	191.9	201.8	265.6	185.4	161.8	192.0	169.3	91.1	174.6	128.4	86.1	86.5	76.1	69.3	73.3	34.5	64.8	19.0			

## Références :

- (1) Préanalyse pour le développement et la modernisation de CELEX. doc IX/F/4-SII/BdD (88) D/288 du 6.12.1980.
- (2) Etude de faisabilité : Modernisation CELEX, système de production des textes (TEXTERFACE) doc IX/I/4-(89) D/194 du 28/6/89 adoptée en réunion le 7/7/89 (doc IX/I/4-(89) D/214 du 14/7/89).
- (3) M. Lloyd, General Access to legal databases/systems in Europe, among others those produced by public authorities in Member States and by Community Institutions, Final Study Report, July 1985, Bruxelles, (chapter II. c : Specific Implications for CELEX, p. 21-25).
- (4) a. Comptes rendus des réunions du groupe informatique juridique du Conseil  
b. Résolution du P.E. sur la simplification, la transparence et la codification du droit communautaire, Débats du P.E. NC 158 p. 386 du 26.6.89 (doc. A2-152/89).  
c. Projet de résolution du P.E. sur le 6ème rapport annuel de la Commission au P.E. sur le contrôle de l'application du droit communautaire - 1988 (doc. A3-158/90 du 13/7/90).
- (5) Questionnaire CELEX, Evaluation, Novembre 1989, document interne CELEX.
- (6) ORACLE, its Architecture, Internals and Database Administration.
- (7) Description générale du système ALICE-T Version 5 (Juillet 90).
- (8) La modernisation CELEX: Aspects Utilisateurs (Note interne IX/F/4-SII (89) D/26 du 27/1/88) notamment l'Annexe III: Les benefices pour l'utilisateur final.

## ANNEXES

0. Table des abréviations
1. Utilisation des bases CELEX
2. Définition des nouvelles bases de diffusion CELEX sous MISTRAL
3. Projet de Manuel CELEX modernisé
4. Fonctionnalités SGBD pour BDD CELEX – Liste de contrôle
5. Structure de la BDP-ARCHIVE
6. Ecrans de saisie BDP-ARCHIVE
7. Calculs de volume
8. L'Organisation des travaux de CELEX après la Modernisation

/u/aleva/MOD/ef2

ANNEXE 0  
TABLE DES ABREVIATIONS

CELEX	Communitatis Europææ Lex
BDD	Bases de Diffusion
ALICE	Alimentation CELEX
ALICE-T	Alimentation CELEX-Textes = TEXTERFACE
ALICE-A	Alimentation CELEX-Analytique
JO	Journal Officiel
TEXTERFACE	TEXT-InterFACE = ALICE-T
TTY	Télétype
ISO	International Standardization Organisation
IC1,2	Imprimerie centrale 1, 2
SZ	Saarbrücke Zeitung
JCL	Job Control Languages
MFTS	Multilateral File Transfer System
CCL	Common Command Language
PME	Petit et Moyen Entreprise
P.E.	Parlement Européen
CES	Comité Economique et Social
OPOCE	Office des Publications Officielles des Communautés Européennes
SGBD	Système de Gestion des Bases de Données
DBM	Data Base Manager
DBA	Data Base Administrator
SGML	Standard Generalised Mark-up Language
SG	Secrétariat Général
DG	Direction Générale
BDP	Base de Production
SQL	Standard Query Language
CJ CJUS	Cour de Justice des C.E.
BDPT-R	Base de production de la partie Textuelle-References
BDPT-T	Base de production de la partie Textuelle-Textes

ANNEXE 1  
UTILISATION DES BASES CELEX  
 (en heures)

ANNÉE	CLX	CEN	CDE	CNL	CIT	CDA	CEL	CES	TOTAL
-------	-----	-----	-----	-----	-----	-----	-----	-----	-------

UTILISATEURS INTERNES

82									4.180
83									5.057
84	6998	382	253						7.633
85	9458	610	315						10.383
86	10927	834	622	130	77				12.513
87	11521	1145	647	257	316				13.886
88	17367	1743	963	380	740	41	19		21.256
89	18783	1978	1071	395	700	347	60		23.337
90*	11014	1398	787	313	385	266	109	22	14.298

UTILISATEURS EXTERNES

82									217
83									439
84									552
85									823
86	688	351	129	44	110				1.322
87	2718	812	517	1125	217				5.389
88	3467	1131	612	1894	502				7.608
89	7215	2577	2526	4936	1215	293			18.765
90*	4401	2821	1670	1434	1223	268	8		11.827

TOTAL

88	20835	2875	1575	2274	1242	41	19		28.864
89	25999	4555	3597	5331	1916	640	61		42.102
90*	15415	4219	2457	1748	1609	535	117	22	26.126

\* Jusqu'au 31/07/90

ANNEXE 2  
DEFINITION DE LA NOUVELLE BASE DE DIFFUSION CELEX SOUS MISTRAL  
 (BASE CLX)

O. LEXIQUES ET CHAMPS ASSOCIÉS

REFER (sans lexique)

BIB

NUMERO = ID ND AN  
 TYPE = DT  
 AUTEUR = AU  
 DESCRIPT = CT CL DE  
 REQUER  
 DEFEND  
 OBSERV  
 INTERV  
 NATIONALA  
 AVOCAT.G  
 JUGE.RAP  
 PROCEDUR  
 CLASSIF = CC  
 LANGUE = LA  
 DESTINA  
 DEPOS  
 RESP (internes)  
 ASSO (internes)

TXT (implicite) = FT UT

TITRE = TI  
 PREAMB VISA  
 CORPUS INDEX  
 FOOTNOTE SOMMAIRE  
 ANNEXE INTITULE  
 MOTIFS  
 DEPENS  
 DISPOSI  
 CONCLUS  
 RESUME = AB  
 TEXT  
 INFO.DIV  
 REF.PUB = SO  
 QUESTION  
 REPOSE  
 DISP.XX

DATE

PUB = PD  
 DOC  
 DEM  
 EFFET  
 FIN.VAL  
 TRANSM  
 SESSION  
 TRANSPOS  
 EFFETMNE  
 ENVOIMNE

REL

BASE  
 CIT  
 CIT.EXT  
 MOD  
 ULT  
 NOTES  
 TRA.PREP  
 TRIB.NAT  
 DISP.NAT

GES (gestion)

REP  
 CREATION = ED  
 UPDATE  
 OPOCE

INFO (sans lexique)

MANUEL

RELAT (sous secret)

BAS.1  
 CIT.1

## A. DESCRIPTION DES CHAMPS ANALYTIQUES

**REFER** sans lexique

Monoséparateur, alphanumérique, mots-clés

Tout document/Secteur

Contient la référence interne et standardisé du document

Longueur max. 15 caractères

Pas de traduction

Structure : SAAATnnnnmmc.ll

S = numéro du secteur

AAA = année (l'année est donné en trois positions; pour tous les documents du 20ème siècle l'année AAA correspond à la combinaison 1AA c.à.d 1er siècle de la vie de CELEX tandis que les documents de l'an 2000 auront comme indicatif 200 ou en général 2AA)

T = classe de type du document

nnnn = numéro du document

mm = numéro d'ordre pour documents avec le même numéro

c = code pour documents spéciaux (p.ex. C pour docs codifiés)

.ll = numéro pour documents-suites

### SECTEURS

1 = Traités

2 = Relations extérieures

3 = Droit Dérivé

4 = Droit complémentaire

5 = Travaux préparatoires

6 = Jurisprudence

7 = Mesures Nationales d'exécution des Directives

8 = Jurisprudence Nationale (en préparation)

9 = Questions Parlementaires

0 = Doctrine (en préparation)

### CLASSES DE TYPES DES DOCUMENTS PAR SECTEUR

Secteur 1

A = Traité EURATOM

B = Traité Adhésion DK IRL GB N

E = Traité CEE

F = Traité de fusion

J = Traité modifications disp budgétaires

G = Traité sur le Groenland

H = Traité Adhésion GR

I = Traité Adhésion ES PO

K = Traité CECA

R = Traité modifications disp financières statut BEI

U = Acte Unique



## Secteur 2

A = Accords

F = Rectificatif

## Secteur 3

R = Règlements

S = Décisions CECA de portée générale

L = Directives

K = Recommandations CECA

D = Décisions

B = Budget

X = Autres actes JO L

Y = Autres actes JO C

C = Censure

F = Rectificatif

## Secteur 4

A = Accords internes

D = Decisions repr. permanents

X = Autres actes JO L

Y = Autres actes JO C

F = Rectificatif

## Secteur 5

P = Propositions Commission

D = Divers actes préparatoires Commission

G = Divers actes préparatoires Conseil et avis conformes CECA

C = Avis PE sur consultation

I = Avis PE sur initiative

B = Avis PE sur le budget

K = Avis Comité Consultatif CECA

S = Avis CES

T = Avis CES initiative

M = Avis Cour des Comptes

## Secteur 6

C = Conclusions

J = Arrêt

O = Ordonnance

S = Saisie arrêt

T = Tierce opposition

V = Avis de la Cour

X = Deliberation

A = Arret TPI  
 B = Ordonnance TPI  
 F = Conclusion devant le TPI  
 D = Tierce opposition devant le TPI

Secteur 7

L = Mesures nationales d'exécution des directives

Secteur 8 p.m.

Secteur 9

E = Question écrite  
 O = Question Orale  
 H = Question heure des questions

Secteur 0 p.m

### RENUMDOCAGE EXTENSIF

Secteur 1:

Les numéros des documents du Secteur 1 sont constitués comme suite:

- Les articles : 1YYTnnnn où YY = année, T = type du traité et nnnn est le numéro de l'article sur 4 positions (voir observations en fin de page);
  - \* Pour les conditions d'adhésion - adaptations, les articles sont numérotés comme Cnnn
  - \* Pour les annexes, les articles sont numérotés comme Ennn
  - \* Pour les protocoles, les articles sont numérotés comme Gnnp
  - \* Pour les échanges de lettres, les articles sont numérotés comme Lnnp
  - \* Pour les conventions d'application ou les conventions sur les dispositions transitoires, les articles sont numérotés comme Pnnp
    - # Les protocoles de ces conventions sont numérotés comme PGnp
  - \* Pour l'acte finale, les articles sont numérotés comme Tnnp
    - # Les annexes de l'acte finale sont numérotés comme TEnnp
    - # Les déclarations associées à l'acte finale sont numérotés comme TGnp
    - # Les articles des procédures d'adoption pendant la période transitoire associées à l'acte finale sont numérotés comme TPnp
  - \* Pour les conventions sur les Institutions communes, les articles sont numérotés comme Vnnp

Observations:

- Les articles nnnnbi, etc.
- Si un annexe nn est citée, elle est citée par nn, article 1 = nn01, article 2 = nn02, etc.

Secteur 3: actes avec extension

Secteur 4: actes KSG ?

Secteur 5:

selon Table 2 (types de documents)

splitage propositions: (ou avec extension (ou avec extension CC) + motifs; plusieurs occurrences de motifs; règlements, directives etc.)

Secteur 6: éliminat

Secteur 7:

mesures nationales d'urgence (document par directives 7aa)

- PC --> P (proposition)
- DC --> D (directive)
- AG --> G (avis de la Commission)
- AP --> C (avis FE)
- IP --> I (avis PE)
- BP --> B (avis CE)
- AK --> K (avis K)
- AC --> E (avis E)
- IC --> O (avis CEE)
- CC --> A (avis C)

REFERS DES DOCS

- INFOCELEX (avec toutes les versions linguistiques) toutes les
- COUVERTURE (avec toutes les versions linguistiques) toutes les
- REPertoire (avec toutes les versions linguistiques) toutes les
- MANUELINTRODUCTIF (avec toutes les versions linguistiques) dans toutes
- MANUELNOV-DE-COM (avec toutes les versions linguistiques) dans toutes

**NUMERO**/BIB

(ND=CCL1, AN=CCL2, ID=SCAD)

Monoséparateur, alphanumérique, mots-clés

Tout document/Secteur

Contient le numéro logique standardisé du document c.à.d. le numéro repris dans le champ REFER sans l'extension pour document suite: SAATnnnn(mm); p.ex. pour la directive 85/85 (CEE) qui a 3 documents suite on aura les documents suivants:

REFER : 3185L0085

NUMERO: 385L0085

REFER : 3185L0085.01

NUMERO: 385L0085

REFER : 3185L0085.02

NUMERO: 385L0085

REFER : 3185L0085.03

NUMERO: 385L0085

Longueur max. 15 caractères

Pas de traduction

A partir de l'an 2000 le champ numero contiendra les données comme suite:

REFER : 3185L0085

NUMERO: 3185L0085 ; 385L0085

etc.

Il y aura peut être nécessité de (re)charger les documents afin qu'ils apparaissent en bonne ordre; dans les titres des documents originaux pour lesquels il existent des documents codifiés il convient d'insérer un message indiquant l'existence et le numero de ces derniers.

**TYPE**/BIB

(DT=CCL)

Monoséparateur, alphanumérique, mots-clés

Tout document/Secteur

Contient le(s) type(s) du document

Longueur max. 1 x 60 caractères (pour quelques langues)  
moyen 35 caractères

Traduction Archive --&gt; Bases

Structure : <type><sup>2</sup>

Synonymie, relation de groupe (pour mots individuels des multitermes), relation d'hierarchie (comme indiquée dans Table 3 pour termes spécifiques).

Re-affectation des codes TYPE selon Table 3; rationalisation données des tables : secrets, 030, 301 (en partie).

TABLE 2: TYPE		
	SYN1	SYN2
TRAITE	1	1
TRAITE EURATOM		1A
TRAITE ADHESION DK, IRL, GB, N		1B
TRAITE CEE		1E
TRAITE FUSION		1F
TRAITE MODI DISP BUDGET		1J
TRAITE GROENLAND		1G
TRAITE ADHESION GR		1H
TRAITE ADHESION ES, PO		1I
TRAITE CECA		1K
TRAITE MODI DISP FINANC STAT BEI		1R
ACTE UNIQUE		1U
ACCORD	A	
ACCORD EXTERNE	2A	
ACCORD BILATERAL		ACC/BIL
ACCORD MULTILATERAL		ACC/MULTI
ACCORD MIXTE		ACC/MIXTE
ACCORD INTERNE	4A	ACCINTERNE
AUTRES ACCORDS		
PROTOCOLE		
ECHANGE LETTRES		ECHLET
CONVENTION		
ARRANGEMENT		ARRANG
AMENDEMENT		
DECLARATION		DECL
RESERVE		RES
MEMORANDUM		
ECHANGE DES NOTES		
CENSURE	3C	
DECISION	3D	
DECISION CECA	3S	DECKSG
DECISION SUI GENERIS		DECBES
DECISION REP	4D	
DIRECTIVE	3L	
REGLEMENT	3R	
REGLEMENT FINANCIER		REGL
REGLEMENT INTERIEUR		REGFINANC
		REGINT
DIVERS ACTES PREPARATOIRES		
DE LA COMMISSION	5DC	
COMMUNICATION		COMMUNIC
PROGRAMME		
RAPPORT		
RAPPORT D'INITIATIVE		IRAPPORT
PROPOSITION DE LA COMMISSION	5PC	
PROJET DECISION		
PROJET DECISION-SUI-GENERIS		
PROJET RECOMMANDATION		
PROJET DIRECTIVE		
PROJET REGLEMENT		
PROJET RESOLUTION		
PROPOSITION DECISION		PDEC
PROPOSITION DECISION SUI GENERIS		PDECB

PROPOSITION DIRECTIVE		PDIR
PROPOSITION REGLEMENT		PREGL
PROPOSITION DIVERSE		
CONCLUSIONS	6C	
CONCLUSIONS DEVANT TPI	6F	
ARRET	6J	
ARRET TPI	6A	
ORDONNANCE	6O	
ORDONNANCE TPI	6B	
SAISIE	6S	SAISIE ARRET
TIERCE OPPOSITION	6T	
TIERCE OPPOSITION DEVANT TPI	6D	
AVIS DE LA COUR	6V	
DELIBERATION	6X	
MESURE NATIONALE D'IMPLEMENTATION		
VERSION CODIFIE		VC
TRAITE CODIFIEE		
TRAITE EURATOM CODIFIE		
TRAITE CEE CODIFIE		
TRAITE CECA CODIFIE		
TRAITE FUSION CODIFIE		
DECISION CODIFIE		
DIRECTIVE CODIFIE		
REGLEMENT CODIFIE		
AUTRE ACTE L CODIFIE		
AUTRE ACTE C CODIFIE		
RECOMMANDATION CECA CODIFIE		
QUESTION PARLEMENTAIRE	9	
QUESTION ECRITE	9E	QUESTE
QUESTION HEURE DES QUESTIONS	9H	QUESTH
QUESTION ORALE	9O	QUESTO
BUDGET	B	
RECOMMANDATION		RECOMM
RECOMMANDATION CECA	3K	RECKSG
RECOMENDATION DE DECISION		RCDEC
RESOLUTION		
RESOLUTION INITIATIVE		IRESOLUT
POSITION COMMUNE		PCC
NON MODIFICATION POSITION COMMUNE		PNM
MODIFICATION POSITION COMMUNE		PMD
REJET POSITION COMMUNE		PRJ
AUTRE ACTE L	X	AUTRE
AUTRE ACTE C	Y	AUTRE
BILAN		
LISTE		
STATUT		
NOTICE		
AVIS		
AVIS CONFORME		AVCON
AVIS D'INITIATIVE		IAVIS
AVIS MODIFIANT		AV-MD
AVIS NON MODIFIANT		AV-NM

AVIS REJETANT	AV-RJ
AVIS SANS RAPPORT	AV-SR
AVIS SOUS RESERVE	AV-RE
AVIS SUR CONSULTATION	

DOCUMENT SUITE	DS
RECTIFICATIF	F
ACTE DE GESTION	3G
INFO	HELP
COUVERTURE = COUV = C	
MANUEL = M	
REPERTOIRE = REP = R	

**AUTEUR**/BIB

(AU=CCL)

Monoséparateur, alphanumérique, (mots-cles)

Tout document/Secteur

Contient l'auteur du document

Longueur max. n x 60 caractères (pour quelques langues) n = ?  
moyen 17 caractères

Traduction archive --> Bases

Structure:

Secteur 1 : <pays ; ><sup>n</sup>

Secteur 2,4 : [<Institution ; ><sup>n</sup> [<pays ; ><sup>n</sup>]

Secteur 3 : <Institution ; ><sup>n</sup>

Secteur 5 : <Institution ; ><Législature ; > [<commission PE ; ><sup>n</sup> [<groupe ; > /AU NOM DU <groupe ; ><nom MEP ; ><sup>n</sup>]

<Institution ; ><section CES ; ><sup>n</sup> <nom MECES ; ><sup>n</sup>

Secteur 6 : <Institution ; > [<nom Avocat Gen ; >]

Secteur 7 : -

Secteur 9 : <Institution ; ><Législature ; > [<groupe ; > <nom MEP ; ><sup>n</sup>]

Rationalisation données des tables 010, 013, 014, 285, 290, 130

Exemples:

<pays>

LES DIX ETATS MEMBRES	10EM
LES ETATS MEMBRES	12EM
LES SIX ETATS MEMBRES	6EM
LES NEUF ETATS MEMBRES	9EM
ACCORD DE CARTHAGENE	ACA
ETATS D'AFRIQUE DES CARAIBES ET DU PACIFIQUE	ACP

AUTRICHE A  
 BELGIQUE B

ISO 3166 (synonymie entre codes à 2 et 3 positions et codes numériques)

Pour assurer une bonne réponse aux questions posées (recall) les descripteurs 6EM, 9EM, 10EM et EM seront suivis des états membres correspondantes; si il y a un 13ème état membre, le descripteur EM sera remplacé avec le descripteur 12EM et à partir de ce moment le descripteur EM sera suivi des noms des 13 états membres.

<Institution>

AGENCE D'APPROVISIONNEMENT EURATOM	AGENCE
ATOMIC ENERGY OF CANADA LIMITED	AECL
AGENCE INTERNATIONALE DE L'ENERGIE ATOMIQUE	AIEA
BUREAU INTERNATIONAL DU TRAVAIL	BIT
BUREAU INTERNATIONAL DES POIDS ET MESURES	BPMN
BUREAU ELARGI	BUREAU-L
CONSEIL D'ASSOCIATION	CASSOC
COURS DE COMPTES DES C.E.	CCCE
COMMISSION CENTRALE POUR LA NAVIGATION DU RHIN	CCR
COMMUNAUTES EUROPEENES	
COMMUNAUTE EUROPEENNE DU CHARBON ET ACIER	CECA
COMMUNAUTE ECONOMIQUE EUROPEENNE	CEE
COMMUNAUTE EUROPEENNE DE L'ENERGIE ATOMIQUE	CEEA/EURATOM
COMITE ECONOMIQUE ET SOCIALE	CES
AUTRICHE-GOUVERNEMENT	
AUTRICHE-PARLEMENT	
...	
CONSEIL	
COMMISSION	

<commission PE>

* Pas de Commission	
COMM. POLITIQUE (73)	73/001
COMM. JURIDIQUE (73)	73/002
....	
....	
COMM. DES PETITIONS (87)	87/018
COMM. TEMPORAIRE POUR LA REUSSITE DE L'ACTE UNIQUE	87/021

<nom MEP>

.... (table 013)

<nom MCES>

.... (table 014)

<nom AVOCAT GENERAL>

.... table 130)

<groupe PE>

GRUPE DES NON INSCRITS	FRACTIO 53/000
GRUPE DES LIBERAUX ET APPARENTES	FRACTIO ..
GRUPE DES SOCIALISTES	FRACTIO ..
...	
...	



<section CES>  
SECTION SPECIALE POUR L'AGRICULTURE ...  
....  
....

Synonymie, relation de groupe

**REQUER**/BIB  
**DEFEND**/BIB  
**OBSERV**/BIB  
**INTERV**/BIB

Monoséparateur, alphanumérique (mots-clés)  
Secteur 6

Contient le(s) nom(s) ou le(s) type(s) du requérant(s), défendant(s), observateur(s), intervenant(s)

Longueur max. 60 caractères (pour quelques langues) n = ?  
moyen 17 caractères

Traduction Archive --> Bases

Structure:

pour REQUER, DEFEND, OBSERV: [<Institution ; >/<pays ;>/<f/p ;>]<sup>n</sup>

pour INTERV : [<Institution>/<pays>/<f/p> -<type interv> ;]<sup>n</sup>

<type interv> = R ou D (auprès du Requerant ou auprès du Defendant)

Synonymie, relation de groupe

A utiliser mêmes tables que AUTEUR; rationalisation tables 110, 120

**NATIONA**/BIB

Monoséparateur, alphanumérique, (mots-clés)  
Secteur 3, 4, 5, 6, 9

Contient la nationalité du destinataire, du requérant, du défendant, de la Cour nationale du membre du CES qui était le rapporteur pour une résolution ou du parlementaire qui était le rapporteur pour une résolution ou qui a posé la question.

Longueur max. n x 80 caractères (pour quelques langues) n=14  
moyen 20 caractères

Traduction Archive --> Bases

Structure <nationalité ;><sup>n</sup>

Synonymie (codes minéralogiques, codes ISO)

Rationalisation tables 140, 290

Exemples

<nationa>	
NATIONALITE ALLEMANDE	D
DIVERSES NATIONALITES	DIV

NATIONALITE PAYS TIERS X  
 NATIONALITE BELGE B  
 ...  
 ...

**AVOCAT.G** BIB  
**JUGE.RAP** BIB

Monoséparateur, alphanumérique, (mots-clés)  
 Secteur 6

Contient le nom de l'avocat général, juge rapporteur

Longueur max. 2 x 60 caractères  
 moyen 17 caractères

Traduction Archive --> Bases

Structure <nom Avocat Gen ;><sup>2</sup>  
 <nom Juge Rap><sup>2</sup>

Relation de groupe pour noms composés

Rationalisation table 130 (cf. AUTEUR)

**PROCEDUR** BIB

Monoséparateur, alphanumérique, (mots-clés)  
 Secteur 6

Contient la nature du recours et éventuellement une précision concernant la solution du litige.

Longueur max n x 80 caractères (pour quelques langues) n=4  
 moyen 70 caractères

Traduction Archive --> Bases

Structure <procedure ;><sup>n</sup> TABLE 100

Synonymie, relation de groupe.

**DESCRIPT** BIB

(CT, DE=CCL, CL=SCAD)  
 Monoséparateur, alphanumérique, (mots-clés)  
 Tout secteur

Contient des mots-clés de classification

Longueur max. n x 80 caractères n=20  
 moyen 200 caractères

Traduction Archive --> Bases

Structure <descript ;><sup>n</sup>

Synonymie, relation de groupe, hiérarchie

Re-affectation codes pour legislation, Jurisprudence sur base de mat, index (thésaurus EUROVOC adapté - remplacement des termes, ameliorations des hiérarchies).

Rationalisation tables 070, 071, 300, 301, 400

**CLASSIF**/BIB

(CC=CCL)

Monoséparateur, alphanumérique, (mots-clés)

Tout secteur

Contient classification thématique sur base du Répertoire ou du Secteur

Longueur max n x 9 caractères n=4  
moyen 33 caractères

Codes numeriques, pas de traduction

Structure: <numéro secteur ;> <code numerique repert ;><sup>n</sup>

Synonymie, relation de groupe

Re-affectation codes Répertoire - Secteurs 5, 6, 9 ?

Rationalisation tables 090, 070, 071

Exemples

<repert>

DISPOSITIONS INSTITUTIONNELLES, GENERALITES	014010
DISPOSITIONS INSTITUTIONNELLES, PARLEMENT	014020
DISPOSITIONS INSTITUTIONNELLES, CONSEIL	014030
.	
.	
.	
CIRCULATION DES MARCHANDISES	02401000
LIBRE CIRCULATION DES MARCHANDISES	02401010
TRANSIT COMMUNAUTAIRE	02401011
AUTRES REGIMES DE LIBRE CIRCULATION	02401012
TURQUIE	02401013
.	
.	
.	
.	

<Secteur>

	SYN1	SYN2
TRAITES	Secteur 1	1
RELATIONS EXTERIEURS	Secteur 2	2
DROIT DERIVE	Secteur 3	3
DROIT COMPLEMENTAIRE	Secteur 4	4
TRAVAUX PREPARATOIRES	Secteur 5	5
JURISPRUDENCE	Secteur 6	6
MESURES NATIONALES D'	Secteur 7	7
EXECUTION DES DIRECTIVES		
JURISPRUDENCE NATIONALE	Secteur 8	8
QUESTIONS PARLEMENTAIRES	Secteur 9	9

DOCTRINE

Secteur 0 0

**LANGUE** BIB

(LA=CCL)

Monoséparateur, alphanumérique, (mots-clés)

Secteurs 1, 2, 3, 4, 6, 7, 8, 10

Contient la langue faisant foi (pour les accords des langues non communautaires), la langue de la procédure ou la langue de publication

Longueur max n x 80 caractères n=?  
moyen 30 caractères

Traduction Archive --&gt; Bases

Structure <langue ;><sup>n</sup>

Synonymie, pour les langues communautaires répétition comme pour AUTEUR; ISO 639

Exemples : table 060**DESTINA** BIB

Monoséparateur, alphanumérique, (mots-clés)

Secteur 3, 4, 5, 9

Contient le destinataire de l'acte ou de l'aide alimentaire

Longueur max ?  
moyen 35 caractères

Traduction Archive --&gt; Bases

Structure

[<Institution ;>]<sup>n</sup>[<pays> [<entité politique ;>]]<sup>n</sup>  
<entité privée, semi-officielle ;>

Synonymie, relation de groupe; pour les EMs répétition des pays comme dans AUTEUR

Rationalisation table 050 (cf AUTEUR)

Exemples entité privée

SHELL CHEMICAL INTERNATIONAL TRADING COMPANY  
INSTITUT NATIONAL DE RECHERCHE AGRONOMIQUE

:

MSG TOTO-PATRIARCHE ....

WINNIE MANDELA

LECH WALESIA

**DEPOS** BIB

Monoséparateur, alphanumérique, (mots-clés)

Secteurs 1, 2

Contient l'organisme depositaire d'un accord.

Longueur max n x 80 caractères n=3  
moyen ?

Traduction Archive --> Bases (pour les entités privées ou semi-officielles avec ou sans traduction)

Structure :  
[<Institution <entité politique> ;>/<pays <entité politique> ;>]<sup>n</sup>

Synonymie, relation de groupe

Rationalisation table 040 (cf AUTEUR)

**RESP** BIB

**ASSO** BIB

Monoséparateur, alphanumérique, (mots-clés)  
Secteurs 3, 4, 5

Contiennent les services responsables et associés.

Longueur max n x 14 caractères n=4 (théoriquement 23)  
moyen 10-30 caractères

Pas de traduction

Structure <DG/DR/DV-AAAA><sup>n</sup> <DG/DR/DV><sup>n</sup>

AAAA (= année) est ajouté quand il y a lieu de  
- restructuration (Commission, interne DG) (affectation par programme sur Archive)  
- changement de l'affectation de l'acte (manuellement sur Archive).

Prevision pour un système d'alerte combinant le fin.val et le service responsable

**PUB** DATE

(PD=CCL)

Monoséparateur, date  
Tout secteur

Contient la date de publication du document

Longueur 10 caractères

Structure AAAA/MM/JJ

**DOC** DATE

Monoséparateur, date avec caractéristiques  
Tout secteur

Contient la date du document avec éventuellement une précision signature ou adoption.

Structure: AAAA/MM/JJ ; AAAA/MM/JJ SIGN ; AAAA/MM/JJ ADOPT

Des renseignements sur le lieu de signature sont à trouver dans INFO.DIV.

**DEM** DATE

Monoséparateur, date  
Secteur 6

Contient la date de la demande à la Cour

Longueur 10 caractères  
Structure AAAA/MM/JJ

**FIN.VAL** DATE

Monoséparateur, date  
Secteurs 1, 2, 3, 4, 5, 7

Contient la date de fin de validité

Longueur 10 caractères  
Structure AAAA/MM/JJ

**EFFET** DATE

Monoséparateur, date avec caractéristiques  
Secteur 1, 2, 3, 4, 5

Contient la date à laquelle l'acte prend effet.

Traduction Archive --> Bases

Structure:

AAAA/MM/JJ EVIG ; AAAA/MM/JJ MISE-EN-APPL ;  
AAAA/MM/JJ PRISE EFFET ;

(cf. table 336)

Fusion champs actuels effet, notif ; des commentaires supplémentaires sur la date EVIG (signature, ratification ou notification) sont à trouver dans info.div.

**TRANSPOS** DATE

Monoséparateur, date avec caractéristiques  
Secteur 3L

Contient la date à laquelle une directive doit être transposée dans les législations nationales.

Traduction Archive --> Bases

Structure:

AAAA/MM/JJ ; AAAA/MM/JJ FR ; AAAA/MM/JJ I ; ....

Prevision de tous les États Membres.

**EFFETMNE** DATE

Monoséparateur, date avec caractéristiques  
Secteur 3L

Contient les dates de prise d'effet des MNEs diverses

Traduction Archive --> Base

Structure:

AAAA/MM/JJ FR1 ; AAAA/MM/JJ FR2 ; ..... FRx ;  
AAAA/MM/JJ GR1 ; ... GRy ; ....

Correspondence avec les données du champ DISP.NAT

**ENVOIMNE** DATE

Monoséparateur, date avec caractéristiques  
Secteur 3L

Contient les dates d'envoi des MNEs à la Commission

Traduction Archive --> Base

Structure:

AAAA/MM/JJ FR1 ; AAAA/MM/JJ FR2 ; ..... FRx ;  
AAAA/MM/JJ GR1 ; ... GRy ; ....

Correspondence avec les données du champ DISP.NAT

**TRANSM** DATE

Monoséparateur, date avec caractéristiques  
Secteur 5, 9

Contient la date d'envoi et/ou de réponse

Traduction Archive --> Base

Structure AAAA/MM/JJ ENVOI ; AAAA/MM/JJ REPONSE

Fusion champs actuels envoi, reponse; des commentaires supplémentaires  
sont à trouver dans info.div.

**SESSION** DATE

Monoséparateur, date avec caractéristiques  
Secteur 5, 9

Contient la date de session et de vote du P.E.

Traduction Archive --> Base

Structure AAAA/MM/JJ DEBAT ; AAAA/MM/JJ VOTE ;

Fusion champs actuels debat, vote; des commentaires supplémentaires sur  
la date SESSION sont à trouver dans info.div.

Observation générale pour les champs REL

Il est proposé de remplacer les parties répétées à chaque début de ligne, par des blancs ou des points (.), tout en tenant chaque ligne comme un paragraphe. Ceci ne permet pas de rechercher par p.ex. :REL (numero &pa ART x) :MOD. Solutions possibles:

- a. pas de remplacement par des blancs ou des points;
- b. introduction d'un NEL (ou d'un LF et CR) à la fin de chaque ligne.

**BASE** REL

Multiséparateur (text, positionnelle)  
Secteur 1, 2, 3, 4, 5, 6, 7

Contient la base juridique de l'acte

Longueur max n x 45 caractères n=30  
moyen 230

Traduction Archive --> Bases

Synonymie

Structure

<NUMERO><sup>n</sup> #  
[<NUMERO> <article> N <alinéa> m #]<sup>n</sup>

Fusion champ actuel traités; Rationalisation tables 020, 370

Exemple :

<numéro>

151E	TRAITE CECA
157E	TRAITE CEE
189H	TRAITE D'ADHESION GRECE
.	.

**CIT** REL

Multiséparateur (text, positionnelle)  
Tout secteur

Contient les citations juridiques d'un acte

Longueur max n x 60 caractères n=30  
moyen 300

Traduction Archive --> Bases

Structure

<numéro><sup>n</sup> #  
[<numéro> <article> N [<alinéa> m] #]<sup>n</sup>  
[<numéro> <article> N [<alinéa> m] <N-><sup>i</sup> <P-><sup>o</sup> #]<sup>n</sup>

Rationalisation table 371



Si il y a lieu de citation d'un document qui ne se trouve pas dans CELEX (règlement de portée limité, document COM publié mais pas dans CELEX etc.) il vaut mieux l'introduire que d'utiliser un numéro fictif.

Si le même numéro est mentionné plusieurs fois la deuxième, la troisième, etc. fois, ils est remplacé par des blancs.

**CIT.EXT** REL

Multiséparateur (text, positionnelle)  
Tout secteur

Contient les citations juridiques d'un acte qui ne font pas partie de la couverture de CELEX

Longueur max n x 60 caractères n=30  
moyen 300

Pas de traduction

Structure  
<ref.publication><sup>n</sup> #

**MOD** REL

Multiséparateur (text, positionnelle)  
Tout secteur

Contient les actes modifiés ou visés par l'acte.

Longueur max n x 80 caractères n=?  
moyen

Traduction Archive --> Bases

Structure  
[<numéro> <mod.active> <part.affectée> n [<DE> date <A> date] #]<sup>n</sup>

Rationalisation tables 275, 372, 375, 395

- Si même <numéro>, même <mod.active>, même <part.affectée> et n différentes, le <numéro> <mod.active> <part.affectée> remplacé par des blancs.
- Données triés sur numéro n.

Exemples :

<mod.active>

ABROGATION  
DEROGATION  
EXTENSION  
CONFIRMATION  
COORDINATION  
PROROGATION  
REPLACEMENT  
SUSPENSION  
APPLICATION DIF.  
ADOPTION

ADOPTION PART.  
 INTERPRETAT.  
 MODIFICATION  
 ADAPTATION  
 SUPPRESSION  
 ADJONCTION  
 COMPLEMENT  
 PROP.DE REMPL.  
 INVALIDATION  
 ANNULATION  
 VALIDATION FR      pour Secteur 6  
 MANQUEMENT EN     pour Secteur 6  
 ...  
 ...  
 VERSION CODIFIE

<partie affectée>  
 DOCUMENT  
 ARTICLE  
 ANNEXE  
 TITRE  
 NUMEROTATION

**ULT**

REL  
 Multiséparateur (text, positionnelle)  
 Tout secteur

Contient les actes qui ont modifié l'acte.

Longueur max n x 80 caractères n=?  
 moyen = ?

Traduction Archive --> Bases

Structure

[<part.affect.> n <mod.passive> <numéro> [<DE> date <A> date] #]ⁿ

Rationalisation tables 270, 375, 396

- Si même <part.affectée>, même n, même <mod.passive> et <numéro> différents, la <part.affectée> n <mod.passive> est remplacé par des blancs.
- Données triés sur n et ensuite sur numéro (à partir de la 2ème position).

Exemples

<partie affectée>

ARTICLE  
 ANNEXE  
 DOCUMENT

<type de modif.>

MODIFIE PAR  
 ABBROGE PAR  
 DEROGAT. PAR  
 EXTENSION PAR

.  
EXTENSION PAR  
. .  
ANNULE PAR  
INVALIDE PAR  
CHANGEMENT DE NUMEROTATION  
VERSION ORIGINALE (pour documents codifiés)

**DISP.NAT** REL

Multiséparateur (text)  
Secteur 3L

Contient les références de publication des mesures nationales d'exécution; traitement comme des textes séparés par version linguistiques (essentiellement pour le grec)

Longueur max. n x 30 n=?  
moyen 360

Traduction Archive --> Bases

## Structure

ETAT-MEMBRE-1 : <commentaire> #  
ETAT-MEMBRE-2 : 1. MNE-1 #  
18 blancs 2. MNE-2 #  
... ..  
18 blancs n. MNE-n #

où:

ETAT-MEMBRE-x : un des douze états membres  
(traduction automatique des codes-pays)

<commentaire> : (traduction automatique)

PAS DE MESURE COMMUNIQUÉ  
MNE PAS NECESSAIRE  
MESURE CONFORME  
MESURE NON CONFORME  
MESURE PARTIELLEMENT CONFORME  
MESURE PAS ENCORE APPRECIÉE PAR LA CCE

Le MNE-n constituée comme suit:

<code.loi> N.xxxx DE date, <code.pub> N.xxx P.yyy DU jj/mm/aa  
(traduction automatique des codes abrégés)

**NOTES** REL

Multiséparateur  
Secteur 6

Contient des articles sur l'affaire

Longueur max. n x 80 n=20  
moyen 320

Pas de traduction

**TRA.PREP** REL

Multiséparateur (text)  
Tout secteur

Contient de l'information sur les travaux préparatoires

Longueur max n x 80 n=20  
moyen 60 caractères

Traduction Archive --> Bases

Structure

[<trav.prep> <ref.pub>]<sup>n</sup>

Validation avec ULT ?

**TRIB.NAT** REL

Multiséparateur, texte  
Secteur 6

Contient des références aux jugements et aux juridictions nationales rendues dans le cadre d'un litige ayant donné lieu à un renvoi préjudiciel.

Longueur max. n x 80 n=20  
moyen 60

Pas de traduction

Structure texte libre

**BAS.1** RELAT

Monoséparateur, alphanumérique  
Secteurs 2, 3, 4, 5, 6, 7

Contient les numéros de la base juridique de l'acte.

Longueur max. n x 17 caractères n=30  
moyen 90

Pas de traduction

Structure <numéro><sup>n</sup>;

**CIT.1** RELAT

Monoséparateur, alphanumérique  
Tout Secteur

Contient les numéros des actes cités.

Longueur max. n x 17 caractères n=30  
moyen 90

Pas de traduction

Structure <numéro><sup>n</sup>;

**REP**/GES

Monoséparateur, alphanumérique  
Secteur 3, 4

Contient de l'information pour la production du Répertoire.

Longueur max. 3 caractères

Pas de traduction

Structure

<rep>  
rep  
lié

**CREATION**/GES

(ED=CCL)  
Monoséparateur, date avec caractéristiques  
Tout document/secteur

Contient la date de création du partie analytique ou du partie textuel  
du document

Longueur max. 10 caractères

Structure AAAA/MM/JJ CRE.A ; AAAA/MM/JJ CRE.T

**UPDATE**/GES

Monoséparateur, date avec caractéristiques  
Tout document/secteur

Contient la date de maj partie analytique  
partie textuelle  
ou 0000/00/00 si le document n'a pas été maj.

Longueur max 10 caractères

Structure AAAA/MM/JJ UPD.A ; AAAA/MM/JJ UPD.T

**OPOCE**/GES

Monoséparateur, alphanumérique  
Tout document/secteur

Contient des informations pour le document delivery à partir des  
systèmes d'archivage de l'OPOCE

Structure:

xxxxxxxxxx-VK ; clé ARCDON; si absent x=0  
yyyyyyyyyy-CC ; numéro catalogue OPOCE/CATEL; si absent y=0  
zzzzzzzzzz-MC ; numéro microfiche OPOCE; si absent z=0  
ppppp-PR1 prix pour envoi normal en ECU  
rrrrr-PR2 prix pour envoi télécopie

ou

VK-xxxxxx ; CC-yyyyy ; MC-zzzzzz ; PR1-pppp ; PR2-rrrrr

## B. DESCRIPTION DES CHAMPS TEXTUELS

### LEXIQUE TXT

(FT, UT=CCL)

Tous les champs sont Multiséparateur, alphanumérique (text)

Tout secteur : TITRE (TI=CCL)

Secteur 2, 3, 4, 5

Secteur 6

PREAMB (5 champs suite)

SOMMAIRE

CORPUS (10 champs suite)

INDEX

FOOTNOTE

INTITULE

ANNEXE (10 champs suite)

MOTIFS (10 ch. suite)

VISA

DEPENS

DISPOSI

CONCLUS (10 ch. suite)

Secteur 5 (COMM) : MOTIFS (10 ch. suite)

Secteur 1, 8, 9, 10 : TEXT (10 ch.suite)

Secteur 5 (PARL), 8, 10 : RESUME (AB=CCL)

Secteur 9 : QUESTION, REPONSE

Dans les titres des documents modifiés indication qu'il y a une version codifié (p.ex. dans le titre de la Directive 370/80 il y aura mention: "Version codifiée 380L0370/c")

### **DISP.XX**/REL

XX = FR, GB, NL, DE, DK, IT, ES, PT, GR, BE, LU, IE (ISO 3166)

Multiséparateur (text)

Secteur 7

Contient le texte intégral des mesures nationales d'exécution, si disponible; traitement comme des textes séparés par version linguistiques (essentiellement pour le grec)

Longueur max. n x 30 n=?  
moyen 360

Traduction/reprise des commentaires

Exemple pour la base française:

DISP.FR : <commentaire> ou texte

DISP.BE : <commentaire> ou texte

POUR LE TEXTE EN NÉERLANDAIS VEUILLEZ CONSULTER LA  
BASE CNL

DISP.LU : <commentaire> ou texte

DISP.UK : VEUILLEZ CONSULTER LA BASE CEN

DISP.IRL: VEUILLEZ CONSULTER LA BASE CEN

DISP.DE : VEUILLEZ CONSULTER LA BASE CDE

DISP.IT : VEUILLEZ CONSULTER LA BASE CIT

DISP.NL : VEUILLEZ CONSULTER LA BASE CNL

DISP.ES : VEUILLEZ CONSULTER LA BASE CES  
 DISP.PT : VEUILLEZ CONSULTER LA BASE CPT  
 DISP.DK : VEUILLEZ CONSULTER LA BASE CDA  
 DISP.GR : VEUILLEZ CONSULTER LA BASE CEL

Pour les documents du secteur 7 les champs DISP.XX contiendront le texte si disponible pour les versions linguistiques correspondantes (p.ex. dans la base CLX les documents du secteur 7 contiendront les champs DISP.FR, DISP.BE, DISP.LU tandis que dans la base CEN ces mêmes documents contiendront les champs DISP.GB et DISP.IE).

Longueur max. illimité

Pour les champs avec champs suite, le non-de-champ est synonyme avec le nom-de-champ01

### **INFO.DIV**/TXT

Multiséparateur, alphanumérique (text)  
 Secteur 2-7

Contient des information diverses sur l'acte.

Longueur max 30  
 moyen ?

Traduction Archive --> Bases

Structure

Secteur

2,3	<validité> <commentaire fin.val>	TABLE 330
2	<commentaire accord>	TABLE 301
2	<signat> <lieu de signat d'un accord>	TABLE 325
3	<transposition> <commentaire transposition directive>	TABLE 361
3	<maj c> <commentaire maj acte>	
3	<effet> <commentaire entrée en vigueur>	TABLE 335
5	<consultation> <commentaire consultation>	TABLE 400
5	<déclar.minor> <commentaire déclaration>	
5	<référence interne>	
6	<affaires jointes> <numéro> <sup>n</sup>	
6	<conclusions jointes> <numéro> <sup>n</sup>	
7	<commentaire secteur 7>	
9	<envoi> <commentaires envoi>	TABLE 340
9	<réponse> <commentaires réponse>	TABLE 350
9	<doc> <commentaires doc>	TABLE 365

Commentaire secteur 7: pour le texte des autres versions linguistiques veuillez voir le même documents dans la version correspondante de CELEX

**REF.PUB**/TXT

(SO=CCL)

Alphanumérique, texte libre, multiséparateur  
 Tout document/secteur

Contient la référence de publication du document.

Longueur max. n x 50 caractères n=9?  
 moyen 50 caractères

Traduction Archive --> Bases

Structure

[<type publ.> nu <du> date <page> pa]<sup>n</sup>  
 [<type publ.> <chapitre> ch <volume> vo <page> pa [<du> date]]<sup>n</sup>

Rationalisation tables 210, 306

Exemples

<type de publication>

JOURNAL OFFICIEL N° L  
 JOURNAL OFFICIEL N° C

RECUEIL DE JURISPRUDENCE N°  
 RECUEIL DE JURISPRUDENCE EDITION ALLEMANDE N°  
 RECUEIL DE JURISPRUDENCE EDITION ITALIENNE N°  
 RECEUIL DE JURISPRUDENCE EDITION NEERLANDAISE N°  
 RECEUIL DE JURISPRUDENCE EDITION ANGLAISE N°  
 RECEUIL DE JURISPRUDENCE EDITION DANOISE N°  
 ...  
 ...

DEBATS DU PE (ED. FRANCAISE) .... N°  
 DEBATS DU PE (ED. ALLEMANDE) .... N°  
 DEBATS DU PE (ED. ITALIENNE) .... N°  
 DEBATS DU PE (ED. NEERLANDAISE) .... N°  
 DEBATS DU PE (ED. ANGLAISE ) .... N°  
 DEBATS DU PE (ED. DANOISE ) .... N°  
 DEBATS DU PE (ED. GRECQUE) .... N°  
 DEBATS DU PE (ED. ESPAGNOLE) .... N°  
 DEBATS DU PE (ED. PORTUGAISE) .... N°

EDITION SPECIALE ANGLAISE  
 EDITION SPECIALE DANOISE  
 EDITION SPECIALE GRECQUE  
 EDITION SPECIALE ESPAGNOLE  
 EDITION SPECIALE PORTUGAISE

.. JC NNNNDMMM-MMM  
 .. UNTS AAAA VOL PPP p. nnn  
 .. UN-04/08/1969-974B  
 .. TRACTATENBLAD 1968 NR 26



**INFO** (sans lexique)  
 Multiséparateur (texte)

Contient des textes informatives non-indexés.

**MANUEL**/MANUEL  
 Monoséparateur, alphanumérique

Contient comme mots-clés les chapitres du manuel (noms des champs, lexiques, secteurs).

Longueur max n x 60 caractères n=3

Structure

<chapitre manuel><sup>n</sup>;

Exemples

<chapitre manuel>

SYNONYMES

MANUEL	CELEX	INTRODUCTION
LEXIQUES		
CHAMPS		
BIB		
NUMERO		
TYPE		
.		
.		
.		
.		

## C. DOCUMENTS SPECIAUX

### 1. Infocellex

REFER : INFOCELEX  
 TYPE : INFO ; HELP  
 INFO :

- > Pour connaître l'état de la base et la dernière MAJ faites COUVERTURE et ensuite m:vi
- > Pour vous aider avec le manuel en-ligne faites MANUEL et ensuite m:vi
- > Pour connaître les codes répertoire faites REPERTOIRE et ensuite m:vi
- > Pour vous aider avec MISTRAL en ligne faites M:SO SO ou M:SO (procédure).

CREATION: YYYY/MM/DD  
 UPD.A : YYYY/MM/DD

### 2. Couverture

REFER : COUVERTURE  
 TYPE : COUVERTURE  
 INFO : L'actuel document info  
 CREATION: ...  
 UPD.A : ...

### 3. Répertoire

REFER : REPERTOIRE  
 TYPE : REPERTOIRE  
 INFO : Le répertoire - table des matières (sur 2 pages max)  
 (pour voir le corpus m:pa txt)  
 CORPUS: Le répertoire - classification complète  
 CREATION: ...  
 UPD.A : ...  
 UPD.T : ...

### 4. Manuel

REFER : MANUEL INTRODUCTION  
 TYPE : MANUEL ; CELEX  
 INFO : Table des matières du manuel  
           CELEX  
           Lexiques  
           Secteur  
           Champs  
           Dates  
           Bibliographie  
           Relations  
           Textes  
           Comment l'utiliser

CREATION: ...  
 UPD.A : ...

REFER : MANUELLEXIQUES  
MANUEL: LEXIQUES  
INFO : Les lexiques de CELEX, comment les utiliser

CREATION: ....  
UPD.A : ...

REFER : MANUELBIB  
MANUEL: BIB  
INFO : Le lexique BIB

CREATION: ...  
UPD.A : ...

REFER : MANUELNUMERO  
MANUEL: NUMERO  
INFO : Le champ numero

#### D. DIVERS

Une Macro lance INFO et m:vi; on visualise le texte du document INFO

E. RELATION ENTRE NOUVEAUX ET ANCIENS CHAMPS

<u>Nouveau champ</u>	<u>Ancien champ</u>
REFER NUMERO	REFER NUM.DOC Renumdocage.
TYPE	TYP.DOC (Table des Secrets) INDEX.CM (301) (en partie) FORME (030)
AUTEUR	AUTEUR (010,013,014) LEGISLAT (285) GROUPE (290) AVOCAT.G (130) (seulement pour les conclusions)
REQUER DEFEND OBSERV INTERV	REQUER (110) DEFEND (120) OBSERV (160)
NATIONA	NATIONA (140) GROUPE (290) (en partie)
AVOCAT.G JUGE.RAP	AVOCAT.G (130) (pour les arrêts) JUGE.RAP (130)
PROCEDUR	PROCEDUR (100)
DESCRIPT	MAT (070,071) MATIERES (070,071) INDEX.CM (301) INFO.DIV (400) DESCRIPT (300) DESC
CLASSIF	TYP.DOC (Table des Secrets) REPERT (090) MAT, MATIERES (070,071)
LANGUE	LANG.FF (060)
DESTINA	DESTINA (050)
DEPOS	DEPOS (040)
RESP ASSO	RESP ASSO
PUB	PUB
DOC	DOC (365) SIGNAT
DEM	DEM

FIN.VAL	FIN.VAL (330)
EFFET	EFFET (335,336) SIGNAT RATIF NOTIF
TRANSPOS	TRANSPOS
EFFETMNE	
ENVOIMNE	
TRANSM	ENVOI REPONSE
SESSION	DEBAT (345) VOTE
BASE	BAS.JUR (370) TRAITE (020)
CIT	CIT.DIV CIT.JUR (371)
CIT.EXT	CIT.JUR
MOD	ACT.MOD (275,375) ACT.VIS (375) ACT.JUR (395,372)
ULT	MOD.ULT (270,375) VIS.ULT (375) ARRETS (396,372)
DISP.NAT	DISP.NAT (nouveau)
NOTES	NOTES (-)
TRA.PREP	TRA.PREP (80, 380)
TRIB.NAT	TRIB.NAT (-)
BAS.1	
CIT.1	
REP	REP
CREATION	CREATION
UPDDATE	UPD.A UPD.T
OPOCE	
DISP.XX	

INFO.DIV

INFO.DIV

REF.PUB

REF.PUB

INFO

INFO

MANUEL

CARACTERISTIQUES DES CHAMPS BDD'S (NEW CELEX)

LEXIQUE CHAMPS	1	2	3	4	5	6	7	8	9	0	IND. POS.	SYN	GRUPE	MONO SEPARATEUR	MULTI SEPARATEUR	HIERA- ROHIE	OBSERVAT.
BIB	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	Ref. unique
REFER	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
NUMERO	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
TYPE	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
AUTEUR	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
REQUIER	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
DEFEND	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
OBSERV	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
INTERV	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
NATIONA	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
AVOCAT.G	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
JUGE.RAP	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
PROCEDUR	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
DESCRIPT	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
CLASSIF	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
LANGUE	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
DESTINA	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
DEPOS	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
RESP	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
ASSO	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
DATE	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date
PUB	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date + carac
DOC	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date
DEM	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date
FIN.MAL	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date
EFFET	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date + carac
TRANSPOS	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date + carac
EFFETME	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date + carac
ENVOIAME	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date + carac
TRANSM	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date + carac
SESSION	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	date + carac
REL.	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
BASE	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
CIT	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
CIT.EXT	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	
MOD	0	0	0	0	0	0	0	0	0	0	-	-	-	X	-	-	

EUROVOC-like

X ?





**DEFINITION DE LA BASE DE PRODUCTION DE CELEX SOUS ORACLE  
(BASE ARCHIVE)**

Champs à ne pas oublier :

**ANALYSTE**

Monoséparateur, alphanumérique  
Tout document/secteur

Contient le(s) nom(s) de l'(des) analyste(s)

Longueur max n x 10 caractères n=2  
moyen

Structure <analyste><sup>n</sup>

**MAJ:INFO**

Multiséparateur, alphanumérique (text)  
Tout document/secteur

Contient des informations sur la maj du document p.ex. n° des caractères, figures et/ou tableaux, information sur l'enrichissement éventuel du texte du document etc.

Longueur max =  
moyen ?

Pas de traduction.

Structure

<maj textes> <n° band maj>  
<commentaires> texte libre

Exemples

<commentaires>  
- champ ult trop long; création document suite;  
- texte provenant de la DG ... contient des erreur à corriger;  
- pas de texte disponible sur support magnétique ....

CHAMPS À GERER COMME CHAMPS TEXTUELS (différents pour chaque version linguistique de la base de diffusion)

Les champs du lexique TXT  
Les champs DISP.XX, NOTES, TRIB.NAT du lexique REL  
Les champs INFO de certains documents

Validations pour chaque champ  
Table champs et lexiques dans différentes versions linguistiques

3 manuels: Manuel BDD, Manuel Analyse, Manuel Gestion/BDP

ANNEXE 3  
PROJET DE MANUEL CELEX MODERNISÉ

À fournir ultérieurement sur demande

ANNEXE 4  
FONCTIONNALITES SGBD pour BDD CELEX – LISTE DE CONTRÔLE

1. INTRODUCTION DES DONNEES

A. DONNEES EN ENTREE

Support physique

- Bandes magnétiques (ISO 1001)
- Transfert de fichier (MFTS-FTAM)

Format logique

- ISO 2709
- SGML

Support des codes 8-bit ou multiple-byte; collating sequence paramétrable par l'utilisateur.

Saisie en ligne avec des écrans

- définis utilisateur
- zones contrôlées selon critères préétablis
- intégration avec système traitement de texte multilingue (8-bit bytes ou multiple-byte)

B. POSSIBILITE DE MAJ

- m.à.j. des données dans les documents
- retrait champs ou documents
- possibilité de m.à.j. en ligne. – temps réduit

C. SPECIFICATIONS

- SGDB multibases avec au moins 25 bases simultanées;
- nombre de documents par base illimité;
- nombre de champs par document au moins 250;
- nombre des caractères par champ illimité pour les champs textuels;
- possibilité d'augmenter les maxima après la création d'une base.

2. INDEXATION

2A. Généralités

- nombre de caractères dans terme indexé > 80
- nombre de termes indexés par champ illimité
- nombre de termes indexés par document illimité
- liste des mots-vides définis par l'utilisateur illimitée
- possibilité de définir :
  - \* des caractères formant des mots (fusion des mots séparés par des tirets)
  - \* des listes de mots avec racine commune ou des règles pour analyse morphologique avant la constitution des fichiers

inverses et pendant l'interrogation (indexer les mots analyse, analyses, analyser, et dérivés par analys... et retrouver tous ces mots en interrogeant avec le mot analysent).

- \* les caractères utilisés dans l'index (correspondance 1 à 2 ou tri ponderé lors de la constitution des indexes)

## 2B. Stratégies d'indexation

- Possibilité d'indexer des expressions entières ou des mots du texte libre selon le champ.
- Possibilité d'intexer en spécifiant
  - \* le champ
  - \* la ligne dans le champ
  - \* l'adjacence c.à.d. indexation positionnelle
- Possibilité de définir entre termes des relations:
  - \* synonymie
  - \* homonymie
  - \* générique et spécifique
  - \* relation de groupe
  - \* autres relations définies par l'utilisateur (jusqu'à 5)
- Possibilité d'introduire des poids aux termes indexés:
  - \* sur base d'une analyse morphologique du texte;
  - \* avec des critères de relevance feedback

## 3. RECHERCHE EN LIGNE

- Disponibilité d'aide en ligne facilement mise à jour.
- Possibilité de créer un fichier d'aide en ligne par base.
- Fichier d'apprentissage.

### A. Recherche et combinaison des termes

- Possibilité d'introduire des recherches multitermes avec affichage des résultats pour chaque terme.
- Possibilité d'afficher l'historique de la recherche. Possibilité de combiner des étapes de recherche entre elles.
- Possibilité de recherche avec des opérateurs BOOLE (ET, OU, SAUF) en général et dans le contexte de l'indexation positionnelle en tenant compte du paragraphe et de la phrase.
- Nombre maximum d'opérateurs dans une étape de recherche =
- Possibilité d'utilisation des parenthèses =
- Troncature
  - \* à droite
  - \* à gauche
  - \* au milieu des mots.
- Masques des mots et proximité
- Possibilité de spécifier le nombre de caractères tronqués.
- Possibilité d'utiliser plusieurs formes de troncature en même temps.
- Possibilité de limiter la recherche à
  - \* un champ
  - \* plusieurs champs définis préalablement (par l'administrateur ou l'utilisateur pendant la recherche)
  - \* étape de recherche précédente
  - \* position des mots
  - \* période(s) de m.à.j. précédente(s)

- Possibilité de sauvegarder les recherches
  - \* temporairement
  - \* en permanence
- Possibilité de ré-exécuter des stratégies de recherche
  - \* entre bases
  - \* entre différents niveaux de sécurité
  - \* après une m.à.j.
- Possibilité d'exécuter des recherches
  - \* pondérées
  - \* en se limitant à certaines dates
  - \* dans un intervalle de nombres, de dates ou de lettres
  - \* en texte libre (p.ex. en cherchant des chaînes de caractères).
  - \* en utilisant les descripteurs communs (expansion de la recherche) entre ensembles de documents.
  - \* en déclarant les relations que le système doit prendre en compte lors des recherches (p.ex. synonymie, hiérarchie) ainsi que leur niveau.
  - \* en déclarant si la recherche doit porter sur les racines des mots indexées ou non.

#### B. Visualisation du fichier inverse

- Possibilité de visualiser des termes du fichier inverse pour un champ donné ou pour un regroupement préétabli des champs.  
Affichage
  - \* du nombre des documents indexés
  - \* des relations entre termes.
- Possibilité de se balayer dans le fichier inverse et sélectionner des documents.

### 4. DONNEES EN SORTIE

#### A. DONNEES IMPRIMEES EN DIFFERE

- Possibilité de définir par l'utilisateur :
  - \* le format d'édition en différé (qui pourra être sauvegardé à l'occasion)
  - \* l'ordre des documents (tri) selon des critères portant sur un ou plusieurs champs.
- Possibilité d'éliminer les documents présents deux fois.
- Possibilité des produire des indexes suivant :
  - \* descripteur dans contexte (Keyword in Context)
  - \* descripteur et contexte (Keyword and Context)
  - \* descripteur hors contexte (Keyword out of Context)

#### B. DONNEES SUR SUPPORT MAGNETIQUE

- Possibilité de décharger les données dans d'autres programmes applicatifs.
- Format de sortie :
  - \* ISO 2709
  - \* SGML

## 5. AUTRES INTERFACES

Nombre maximum de termes dans un profil

Possibilité d'établir des profils avec :

- logique Boole
- recherches pondérées

Possibilité d'inclure les relations entre termes

Possibilité de modifier des profils :

- ajouts nouveaux termes
- éliminations de termes

Possibilité de définir l'exécution des profils

- à des intervalles prédéfinis par l'utilisateur
- à la demande de l'utilisateur

- automatiquement quand on ajoute des documents nouveaux

Possibilité de définir le format de sortie des profils.

Possibilité de (télé)décharger les données à une autre machine/progiciel pour permettre la manipulation/reformatage.

## 6. SECURITE

Possibilité de définir plusieurs niveaux de mots de passe pour un utilisateur

Possibilité de réduire les niveaux d'accès

à la lecture sans possibilité de modification de documents

à certaines bases

à certains documents d'une base

à certains champs d'un document

Possibilité d'avertir ou non l'utilisateur que ces facilités/documents ne sont pas disponibles.

## 7. MAINTENANCE

ANNEXE 5  
STRUCTURE DE LA BASE DE PRODUCTION ARCHIVE (BDPA)

0. Introduction :

L'objectif de ce document est de fournir un premier aperçu de la structure relationnelle de la BDP-Archive par les secteurs et les champs les plus importants. Ce document doit être validé et consolidé au niveau de la préparation des spécifications par l'appel d'offre, mais le cadre général est correct.

1. Structure générale :

Le BDP-Archive aura comme fonction le stockage permanent et dans une forme codifiée de toutes les parties bibliographiques des documents CELEX destinés à être chargés dans les BDP's.

Le fait que la plupart du contenu de la BDP-Archive soit codifié est destiné à permettre que celle-ci agisse comme "moteur de production" des versions linguistiques.

Le BDP-Archive n'est pas orienté vers la recherche documentaire, et tous les accès à l'information seront définis à l'aide des clés et en utilisant les facilités fournies par le logiciel.

La structure de BDP-Archive, à caractère relationnel, est composée d'un groupe de tables organisées de la façon suivante :

- . Tables associées aux documents, avec la partie analytique d'un document CELEX dans une forme codifiée est indépendante de toutes versions linguistiques.

Elles sont organisées en :

- a) Tables générales contenant les attributs communs à tous les secteurs.
- b) Tables par secteur. A chaque secteur de CELEX correspondra un groupe de tables avec les attributs spécifiques pour ce secteur.
- . Tables de validation des différents codes utilisés organisés dans CELEX avec une structure de thésaurus simple si nécessaire.
- . Tables de traduction avec la structure générale suivante :  
TABLE-TR (code, langue, description-langue).
- . Tables de relation entre les documents.

## II. TABLES-DOCUMENTS

## II.1 Tables générales

- 1) Tables DOC-G (document général). La table contiendra tous les attributs d'un document qui sont en relation 1:1 avec celui-ci.

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Refer (clé)	X(15)	Refer document (voir Refer annexe 2)
Numéro-1 (clé-2)	X(12)	Numéro CELEX 1 (Voir Numéro annexe 2)
Numéro-2	X(12)	Numéro CELEX 2 (Voir Numéro annexe 2)
Secteur	A	Numéro secteur
C-type-1	X(2)	Type document, validation T-TYPE
C-type-2	X(2)	Validation T-TYPE
Date-pub	date	date publication
Date-doc	date	date document (voir champ DOC annexe 2)
Date-sig	date	date signature (voir champ DOC annexe 2)
Date-adopt	date	date adopt (voir champ DOC annexe 2)
Date-crea	date	date création partie analyse
Date-cre-t	date	date création partie texte
Date-upd-a	date	date maj partie analyse
Date-upd-t	date	date maj partie texte
Arcdon	X(10)	Clé base ARCDON
Catel	X(10)	catalogue OPOCE/CATEL
Microfiche	X(10)	numéro microfiche

- 2) DESCRIPT (Relation n:m avec un document)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Clé {numéro	X(12)	Refer document
└C-desc	X(10)	Code descripteur, validation T-DESCRIPT

- 3) CLASSIF (Relation N:M avec un document)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Clé {numéro	X(12)	Refer document
└C-classif	X(10)	



## 11.2 Tables documents secteur 1

### 1) Table DOC-S1

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	numéro CELEX
Date-fval	date	date de fin de validité
Date-evig	date	
Date-mep	date	prise effet
Date-pe	date	mise en application

### 2) Table AUT-PAYS (voir tables communes)

### 3) Table LANGUE (voir tables communes)

### 4) Tables DEPOT

#### 1) DEPOT-I

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	Numéro CELEX
C-inst	X(4)	validation contre table T-PAYS
C-ent-administr.	X(5)	validation entre T-entité-administrative

#### 2) DEPOT-P

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	Numéro CELEX
C-pays	X(4)	validation contre table T-pays
C-ent-administ.	x(5)	validation T-entité-administrative

## 11.3 Tables des documents secteur 2

### 1) Table DOC S2

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numero	X(12)	numéro CELEX
Date-fval	date	date fin de validité
Date-evig	date	
Date-mep	date	] voir champ DOC (annexe 2)
Date-pe	date	

- 2) Table AUT-PAYS (voir tables communes)
- 3) Table AUT-INST (voir tables communes)
- 4) Tables LANGUE (voir tables communes)
- 5) Tables DEPOT (Idem doc. secteur 1)

#### 11.4 Tables documents secteur 3

##### 1) Table DOC-S3

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	numéro CELEX
Date-fval	date	
Date-evig	date	
Date-mep	date	
Date-pe	date	] voir champ DOC (annexe 2)

- 2) Table AUT-PAYS (voir tables communes)
- 3) Table AUT-INST (voir tables communes)
- 4) Table LANGUE (voir tables communes)
- 5) Table NATIONALITE (voir tables communes)
- 6) Tables destinataire

##### 1) DESTINA-I

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
<u>Numéro</u>	X(12)	numéro CELEX
<u>C-instit</u>	X(4)	code institution, validité centre T-INST

##### 2) DESTINA-P

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
<u>Numéro</u>	X(12)	numéro
<u>C-pays</u>	X(4)	code pays, valider contre T- PAYS
C-entité-politique	X(4)	code entité politique, contre T-ENTITE-POLITIQUE

7) Tables RESP

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	Numéro CELEX
C-unité	X(10)	Code unité administrative, valider contre T- DIRECTORY

8) Table ASSO

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	Numéro CELEX
C-unité	X(10)	Code unité administrative, valider contre T- DIRECTORY

9) Table date de transposition directive (TRANSP)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(11)	Numéro CELEX
c-langue ] clé	A(2)	
Date-transpos	date	

10) Prise d'effet MNE

<u>Champs</u>	<u>Types</u>
Numéro	X(11)
c-langue	A(2)
N°-mne	G(2)
date-effetmne	date

11) Envoi MVE

<u>Champs</u>	<u>Types</u>
Numéro	X(12)
C-langue ] langue	
No-mne	
Date-envmne	9(2) date

11.5 Tables documents secteur 4

1) Tables DOC-S4

<u>Champs</u>	<u>Types</u>	
Numéro	X(12)	
Date-fval	date	
Date-evig	date	voir champs DOC (annexe 2)
Date-map	date	
Date-pe	date	

2) Table AUT-PAYS (voir tables communes)

3) Table AUT-INST (voir tables communes)

4) Table LANGUE (voir tables communes)

5) Table NATIONALITE (voir tables communes)

6) Table DESTINATAIRE (voir réf. 11.4 secteur 3)

4) Table RESP,ASSO (voir secteur 3 point 11.4)

11.6 Tables documents secteur 5

1) Table DOC-S5

<u>Champs</u>	<u>Types</u>	<u>Description</u>
Numéro(clé)	X(12)	numéro CELEX
C-instit	X(4)	codification, validation contre T-INSTIT
C-legis	X(3)	code législative, validation T-LEGIS
Date-fval	date	voir annexe II, champ DATE
Date-evig	date	
date-map	date	
Date-pe	date	
Date-envoi	date	
Date-reponse	date	
date-débat	date	
Date-vote	date	

2) Table AUT-SCES (section CES)

<u>Champs</u>	<u>Types</u>	
Numero clé	X(12)	
C-SCES clé	X(5)	validation table T-SEC-CES

3) Table AUT-MCES (membre CES)

<u>Champs</u>	<u>Types</u>
Numéro clé	X(12)
C-mem-ces clé	X(5) validation table T-MEM-CES

4) Table AUT-CPE (commission PE)

<u>Champs</u>	<u>Types</u>
Numéro clé	X(12)
C-com-pe clé	X(5) validation autre table T-COM-PE

5) Table NATIONALITE (voir tables communes)6) Table AUT-GRP (groupes politiques)

<u>Champs</u>	<u>Types</u>
Numéro	X(12)
C-groupe	X( ) validation autre T-GROUP

7) Table AUT-MPEC (membres du PE)

<u>Champs</u>	<u>Types</u>
Numéro	X(12)
C-mem-pe	X(10) validation table T-MEM-PE

## 11.7 Tables documents secteur 6 (à compléter)

## 11.8 Tables documents secteur 7 (à compléter)

## 11.9 Tables documents secteur 9 (à compléter)

Tables communes à plusieurs secteurs :

Ces tables sont le résultat d'optimiser des tables avec un contenu identique au niveau des différents secteurs. Généralement, elles sont en relation N:M avec le document.

1) Table AUTEUR-PAYS

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	numéro CELEX
C-pays	X(4)	code pays (validation T-PAYS)

2) Table LANGUE

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
Numéro	X(12)	numéro CELEX
C-langue	X(2)	code langue (validation T-LANGUE)

## III Tables de validation

1) Table des types de documents (T-TYPE)

Contenu : les différents types de documents ainsi que les secteurs dans CELEX. Cette table groupe les concepts de secteur, type et classe de document.

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
C-type	X(2)	code type document. Pour le type de document est identifié par le couple (n°-secteur, c-type)
Secteur	9	numéro secteur, valeur entre 0 et 9
C-classe	A	code type associé à la classe de type de document qui fait partie du numéro CELEX
C-pere	X(2)	couple (n°-sect, c-type) qui constitue le pere de cette occ? dans la hiérarchie entre types.

Les secteurs, isolement, sont identifiés par la clé (n° secteur, c-type) avec c-type=n°-secteur.

NB : On peut utiliser le numéro de secteur comme une partie du code du type du document.

2) Table des descripteurs (T-DESCRIPT)

Contenu : la liste des codes descripteurs à utiliser dans CELEX. Ils seront assimilés à ceux d'EUROVOC.

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
<u>C-desc</u>	X(10)	code descripteur (on utilisera les code EUROVOC)
<u>C-desc.p</u>	X(10)	code du descripteur père dans la hiérarchie (les racines auront la valeur (0))

- Associé à une table de traduction
- Exemple : une hiérarchie suivante

```

A
  A1
  A2
    A21
    A22
  A3
    
```

Sont représentés par

<u>c-desc</u>	<u>c-desc-0</u>
A	0
A1	A
A2	A
A21	A2
A22	A2
A3	A

### 3) Table de classification thématique T-CLASSIF

Contenu : liste des copies de classification thématique

<u>Champ</u>	<u>Type</u>	<u>Description</u>
<u>C-class</u>	9(9)	code de classification thématique

Associée à une table de traduction

### 4) Table de classes de documents (T-CLASSE)

Contien : classe des types de documents CELEX pour la construction du numéro CELEX

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
<u>N-secteur(clé)</u>	9	numéro secteur
<u>C-classe(clé)</u>	A	code classe
<u>D-classe</u>	X(50)	description classe

5) Tables généralisées des pays (T-PAYS)

Contenu : liste des codes pays dans le sens général du terme

<u>Champ</u>	<u>Types</u>	<u>Description</u>
C-pays	x(4)	code associé au pays. Dans la plupart des cas le code est la plaque minéralogique du pays. Dans d'autres cas ça peut être un groupement de pays par exemple (EM, 9EM, 6EM)

6) Table généralisée des Institutions (T-INST)

Contenu : liste des codes des Institutions diverses. Le terme institution est employé dans un sens large.

<u>Champ</u>	<u>Type</u>	<u>Description</u>
C-instit	X(4)	code institution

7) Table des codes de législation (T-LEGIS)

Contenu : liste des codes associés aux législations du parlement

<u>Champ</u>	<u>Type</u>	<u>Description</u>
C-legis	X(3)	code utilisé pour identifier la législature.

8) Table des codes de groupe (T-GROUP)

Contenu liste des codes dans groupes politiques de PE

<u>Champ</u>	<u>Type</u>	<u>Description</u>
C-groupe	X(10)	Code d'identification de groupe

- Construction du code du groupe
  - . sigle du groupe
  - . année
  - . législation

A l'heure actuelle la clé est FRACTION-AA/<no-sequence>, à chaque nouvelle législature il faudra introduire les groupes parlementaires



9) Tables membres de PE (T-MEMPE)

Contenu : liste des membres du PE

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
C-mem-PE	X (5)	code associé au une personne
C-legisl-PE	X (3)	code législation
nom	X (120)	nom personne
C-national	X (2)	code nationalité

- L'année de nomination comme membre du P.E. peut faire partie de la clé C-mem-PE
- A chaque nouvelle législature, il faudra introduire les noms des parlementaires

10) Table commission PE (T-COM-PE)

Contenu :  
Liste des codes des commissions du Parlement Européen

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-comm-pe	X(5)	Code commission
s-comm	X(90)	Synonymie
c-legis	X(3)	Code legislation

- A l'heure actuelle la clé est construite comme AA/<no-sequ> avec l'année <AA> suivie du numéro séquentiel.

11) Table section CES (T-SEC-CES)

Contenu :  
Liste des codes des sections du Comité Economique et Social.

<u>Champ</u>	<u>Type</u>	<u>Description</u>
c-sec-ces	X(5)	clé

- Construction de la clé : CES\_<AA>/no\_séquence> où <AA> est l'année.

12) Table des membre de CES (T-MEM-CES)

Contenu :  
Liste des membres du Comité Economique et Social

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-mem-ces	X (5)	Code d'identification
nom	X(120)	Nom du membre
année	X (4)	Année

- Construction clé :

13) Table des avocats de la CJ (T AVOCAT)

Contenu :

Liste des avocats généraux de la Cour de Justice.

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-avocat	X ( )	
année	X (4)	Année de nomination
nom-avocat	X (120)	Nom de l'avocat

- Construction de la clé :

14) Table des codes de nationalité (T NATIONALITE)

Contenu :

Liste des nationalités dans un sens général.

<u>Champ</u>	<u>Type</u>	<u>Description</u>
c-natio	X(4)	Code de la nationalité

15) Table des juges rapporteurs de la CJ (T-JUGE-RAP)

Contenu :

Liste des juges rapporteurs de la Cour de Justice.

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-juge	X (4)	Identification du juge
année	X (4)	
nom-juge	X 120)	Nom du juge

16) Table des codes procédures de la CJ (T-PROCEDURE)

Contenu :

Liste des procédures de la Cour de Justice.

<u>Champ</u>	<u>Type</u>	<u>Description</u>
c-procédure	X (5)	clé Code d'identification

17) Tables des langues (T-LANGUES)

T-LANGUES

Contenu :  
Liste des langues.

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-langue	X (4)	Code langue. On utilisera les normes ISO tant que possible
s1-langue	X (20)	Synonymie
s2-langue	X (20)	Synonymie
d-langue	X (70)	Nom langue

18) T-ENTITE-POLITIQUE

Contenu :

<u>Champ</u>	<u>Type</u>	<u>Description</u>
c-ent-pe	X(5)	

19) T-ENTITE-PRIVEE

Contenu :

<u>Champ</u>	<u>Type</u>	<u>Description</u>
C-ent-prive	X(5)	

20) T-DIRECTORY

Contenu :  
Contient, sous forme codifiée, les diverses unités de l'administration de la Commission.

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-unité	X(10)	Code identification de l'unité
année	X(4)	

- Construction de c-unité sur la base du niveau hiérarchique et de l'année de création.

Observation

- Certaines des tables citées au-dessus pourront lors de l'implantation être groupées dans une seule table en y ajoutant un attribut pour identifier le type de contenu.

#### IV. Table de traduction

##### 1) Traduction des types et secteurs (TR-TYPE)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
no-sect(clé)	9	identification type/secteur
c-type(clé)	X(2)	
c-langue(clé)	A (2)	code langue (une des langues communautaires)
syn-1	X (20)	synonymie 1
syn-2	X (20)	synonymie 2
d-type	X (70)	description type

##### 2) Traduction des descripteurs (TR-DESC)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-desc(clé)	X(10)	code descripteur
c-langue	A(2)	code langue
d-descr	X(70)	description-descripteur

On utilise les titres EUROPE

##### 3) Traduction de classification thématique (TR-class)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-class(clé)	G(10)	code classification
c-?(clé)	A(2)	code langue
d-class	X(70)	descripteur classification

##### 4) Table de traduction dans PAYS (TR-PAYS)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-pays (clé)	X(4)	code pays
c-langue (clé)	X(2)	code langue
d-pays	X(70)	descripteur de pays dans la langue

##### 5) Table de traduction des Institutions (TR-INST)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-inst (clé)	X(4)	code institution
c-langue (clé)	A(2)	code langue
d-instit	X(70)	description de l'institution dans la langue

6) Table de traduction des législations (TR-LEGIS)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-legis (clé)	X (3)	code identification législative
c-langue (clé)	A (2)	code langue
d-legis	X (70)	traduction de la législation dans la langue

7) Table des traduction des groupes (TR-GROUP)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-groupe (clé)	X( )	code groupe
c-langue (clé)	X(2)	code langue
synonymie	X(40)	synonymie pour cette langue
d-groupe	X(40)	nom groupe dans langue

8) Table des noms des commissions de PE (TR-COM-PE)

Contenu : liste des noms dans les différentes langues des commission du PE

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-com-pe (clé)	X(5)	code commission
c-langue (clé)	A(2)	code langue
synonymie	X(70)	synonymie
d-com-pe	X(70)	nom commission

9) Table des cours de section du CES (TR-SEC)

Contenu : liste des noms des sections du Comité Economique et Social

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-sec-ces (clé)	X(5)	code section comité économique et social
c-langue (clé)	A(2)	
synonymie	X(70)	
d-sec-ces	X(70)	description section

10) Codes de nationalité (TR-NATIONALITE)

Contenu : listes des nationalités utilisés

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-nation (clé)	X (4)	code nationalité
c-langue (clé)	A (2)	
synonymie	X (20)	
c-nation	X (70)	

11) Table des noms des procédures (TR-PROCEDURES)

contenu : liste des noms des procédures de la CJ

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-procédure	X (5)	code procédure
c-langue (clé)	X (2)	
d-procédure	X (70)	description procédure dans la langue

12) Liste des entités politiques (TR-ENTITE-POLITIQUE)

13) Liste des entités privées (TR-ENTITE-PRIVE)

14) Liste des unités administratives (TR-DIRECTORIE)

Contenu : liste des unités administratives de la Commission

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
c-unité (clé)	X(10)	code d'identification unité
c-langue(clé)	A(2)	code langue
d-unité	X(70)	nom de l'unité

Observation : lors de l'implantation, certaines tables avec une structure commune pourront être groupées

V. Table des relations entre documents

1) Base juridique (T-BASE)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
numéro(clé)	X(12)	
n.ordre(clé)	9(2)	numéro d'ordre
base clé	X12)	numéro celex qui constitue la base juridique
article	9(3)	
alinéa	9(2)	

2) Citations (T-CIT)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
numéro(clé)	X(12)	
n.ordre(clé)	G(2)	numéro d'ordre
cit (clé)	X(12)	numéro de citation
article	Q(3)	
alinéa	Q(2)	
page	X(20)	
attendu	X(20)	

3) Citations externes (T-CITEXT)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
numéro	X(12)	
n.ordre	G(2)	numéro d'ordre
ref.pub	X(800)	

4) Actes modifiés (T-MOD)

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
numéro(clé)	X(12)	
n.ordre (clé)	G(2)	
num-mod	X(12)	acte modifié
c-mod.	X(3)	validation T-MODF
c-partie	X(1)	validation T-PARTIE
date-i	date	
date-f	date	

5) T-ULT

<u>Champs</u>	<u>Types</u>	<u>Descriptions</u>
numéro(clé)	X(12)	
n.ordre (clé)	G(2)	
acte-mad	X(12)	numéro acte modifié
c-mod	X(3)	validation T-MODIF
c-partie	X(1)	validation T-PARTIE
date-i		
date-f		

ANNEXE 6

ECRANS DE SAISIE BDP-ARCHIVE

À fournir ultérieurement



ANNEXE 7  
CALCULS DE VOLUME

Pour le calcul de volume des bases BDP-Textes et BDP-Archive il a été utilisé une simplification de la formule de calcul fournie par ORACLE (v. ref. 6 et Annexe 7).

DATA :

$$BL = LL + NC \times 2 + 4$$

$$NO \times BL$$

$$BLK = \frac{\quad}{1024 - 76} + 2$$

où :

- BL est le nombre de bytes par ligne
- LL est la longueur d'une ligne en bytes
- NC est le nombre de colonnes dans une table
- BLK est le nombre des xxxxxx logiques en Kbytes
- NO est le nombre des occurrences d'une table

INDEX :

$$BLK = \frac{1.1 \times NO \times LI}{(1024 - 76)}$$

où :

- LI est la longueur d'un index

1. Systeme local :

a) Base de production des textes (ALICE-T) :

La BDPT (cf. [5]) est structurée en deux parties :

- . la BDPT-R, base de référence type relationnelle implémentée en ORACLE et composée des trois tables : TXT\_PROV (contenant les documents en production), TXT\_HISTO (historique du chargement) et PRESENCE (image des BDD's (aussi valable pour ALICE-A))
- . la BDPT-T composée des fichiers séquentiels avec les textes des documents qui sont référencés dans la BDPT-R.

Estimation de volume :

1. BDPT-R :

Table	LL	NC	NO	LI-1	LI-2	BLK_DATA (MB)	BLK_INDEX (MB)
TXT_PROV	342	22	10.000	6	19	4.095	0.290
TXT_HISTO	61	10	90.000	21	-	8.072	2.193
PRESENCE	21	4	1.260.000	17	-	41.203	24.854
						53.370	27.337

TOTAL BDPT\_R = 80.707 +/- 81 MB

2. BDPT-T :

L'espace correspondant à la BDPT est géré comme un dépôt temporaire des textes pour les documents qui sont en attente de chargement dans les BDD's. Il est prévu un volume total de l'ordre de 100 MB, ce qui permet un stockage temporaire de plus de 10.000 documents.

Croissance annuelle :

BDPT-R			
TXT_PROV	nulle		
TXT_HISTO	8	MB/an	(90.000
		documents/an)	
PRESENCE	4'7	MB/an	
BDPT-T			
	nulle		

En conséquence on a pour la BDPT :

- \* volume initial : 181 MB
- \* croissance annuelle : 12'7 MB

b) BDP-Archive :

Le calcul pour l'estimation de volume est dans ce cas moins précis que pour la partie texte étant donné qu'on ne dispose encore pas de la structure finale (voir annexe 3). Cependant, pour simplifier, on va grouper les différentes tables en 3 groupes et faire des hypothèses "raisonnables" sur les valeurs nécessaires pour faire l'estimation.

Volume initial :

Table	LL	NC	NO	LI-1	BLK_DATA (MB)	BLK_INDEX (MB)
DOCUMENT	700	40	140.000	20	115.781	3.249
VALIDATION/ TRADUCTION	1000	10	90.000	10	97.215	1.045
RELATION	70	5	280.000	20	24.810	6.498
					237.806	10.792

TOTAL : 258.598 +/- 259 MB

Croissance annuelle :

DOCUMENT (10.000 doc./an)	(8.04 MB)
VALIDATION/TRADUCTION	négligeable
RELATION (30.000 occurrences)	(3.36 MB)
	TOTAL : 8.40 MB

En conséquence on a pour la BDP-Archive :

- Volume initial : 259 MB
- Croissance annuelle : 8'5 MB

## c) Gestion :

Pour la gestion de l'application en local on va supposer un espace de gestion de l'ordre de 300 MB.

Volume total système local :

BDP-Textes :	181
BDP-Archive :	259
Gestion :	300
	—

TOTAL : 740 MB

ANNEXE 8

L'Organisation des travaux de CELEX après la Modernisation



DIRECTION GÉNÉRALE  
PERSONNEL ET ADMINISTRATION  
DIRECTION INFORMATIQUE  
BASES DOCUMENTAIRES  
Celex

(91) D/165

L'ORGANISATION DES TRAVAUX DE CELEX  
Les besoins vs. les ressources disponibles

Actions à caractère verticale

<u>Fonction</u>	<u>Description des tâches</u>	<u>Grade</u>	<u>Situation</u>
<u>LE DATA BASE MANAGER</u>	Management; établit des objectifs stratégiques; relations avec CIC, GIJC, autres DGs.	A 3/4/5 Juriste	Mme Bensch ??
Sécretariat	Travaux de secrétariat	2 C	Mme Van Impe ??
1. <u>PRODUCTION DONNÉES</u>	Supervision globale de la production des données, (analyse, saisie, m.à.j., codification, qualité données)	1 A Manager	M. Dewez
1a <u>Analyse</u>	Methodologie et contrôle de l'analyse; supervision de l'équipe des analystes	1 A Juriste	??
Équipe analystes	Effectuer l'analyse de jour en jour	s/t 7 Juriste	MI Bartali M. Lelievre MI Martin MI De Coster M. Da Silva ?? ??
Répertoire	Suivi de la production du Répertoire (édité 2 x par an); corrections; rel. avec OPOCE	1 A Juriste	??
1b <u>Codification</u>	Coordination des travaux de codification des actes de la législation; suivi des actions; coordination de l'équipe des chargés de la codification (v. actions horizontales) et éventuellement des DGs	1 A Juriste	Mme Devos ??

2. <u>PRODUCTION BASES</u>	Gestion de la production des bases, suivi des māj en local et en central, supervision de la gestion informatique en local et central	1 A Manager	??
2a <u>Coordination</u> <u>Encodage -</u> <u>Alimentation</u>	Coordination encodage; gestion contrats firmes d'encodage; gestion données en provenance des autres Institutions et de l'OPOCE	1 B	Mme Hubert
Equipe encodage	Saisie des données	4 C	Mme Gysels Mme Geiragat Mme Swinnen Mme Dalamaga
Dispatching, Gestion JOs, Numdocage, Validations num.doc's	Dispatching, Gestion JOs, Numdocage des documents à analyser, Validations des num.doc's dans TEXTERFACE	1 B/C 1 D	Mme Beviere ?? M. Maleve
2b <u>Gestion</u>	Gestion Informatique du système CELEX (local et central); suivi et coordination des travaux; intervention pour faire avancer les choses vu l'instabilité et la complexité du système	1½ B Informa ticien	??
Gestion ALICE	Gestion Informatique et gestion des données pour le système local de production de CELEX; chargement des données sur le système local; donne du travail à l'équipe lb; back-ups de données; dechargement des données vers le système central; reception nouvelles versions du logiciel applicatif; en deux mots faire "tourner" le système local ALICE	1½ B/C Informa ticien	M. Albanese
LSA	Suivi du système local; fonctions LSA	1 B Informa ticien	IUC ??

Gestion BDDs	Reception des fichiers transmis par la gestion ALICE; lancement des m.à.j. en central; suivi des bases de diffusion - reorganisations; travaux d'exploitation en batch	2 B/C Informaticien	Gestion CC ??
3. <u>DIFFUSION</u>	Coordination des actions concernant les activités suivantes: Formation; Documentation; Promotion; avec l'aide des personnes ayant les connaissances linguistiques nécessaires (v actions horizontales); formation formateurs; suivi sous-traitance de la formation; suivi promotion avec OPOCE, EUROBASES	1 A Manager	M. Krafft ??
Serveurs	Suivi des serveurs; lancement des profils pour les livraisons initiales et les m.à.j. maintenance du Manuel Serveurs; (il peut appartenir à l'équipe Gestion 1d); relations avec l'OPOCE et les Serveurs; cours de formation des Serveurs; helpdesk pour les Serveurs	1 B Informaticien	??

Actions à caractère horizontale Production-Diffusion

<u>Fonction</u>	<u>Description des tâches</u>	<u>Grade</u>	<u>Situation</u>
<u>QUALITÉ DES DONNÉES</u>	Suivi du vocabulaire contrôlé par langue (9 langues officielles); Suivi de la qualité des textes par langue (splittage des textes, corrections fautes)	9 x A/LA/B Connais sances linguis tiques	
<u>FORMATION</u>	Assurer la formation pour le système dans les 9 langues officielles; éventuellement assistance à la promotion du système	9 x A/LA/B Connais sances linguis tiques	
<u>CODIFICATION DOCUMENTATION</u>	- Produire et ensuite maintenir les versions codifiées des actes de la législation. Rédiger et puis maintenir la documentation pour les 9 versions de CELEX (Manuel, Support de cours, Vade-Mecum, Exercises)	9 x A/LA/B Connais sances linguis tiques	

Pour tous les taches horizontales l'on dispose actuellement les ressources suivantes par langue:

FR -  
EN M. Massie (qui suit aussi la m.à.j.) et 50% de Mme Senez\*  
DE 50% de Mme Arnold\*  
IT Mme De Persio (qui suit aussi un peu le Répertoire)  
NL -  
DA 50% de M. Bock\*  
GR M. Alevantis (qui suit aussi la Modernisation) et M. Meimaridis  
ES 50% de M. Caton\*  
PO 75% de M. Pereira\* (qui s'occupe actuellement de la creation de la base portugaise)

Les personnes marquées avec \* ne faisant pas partie intégrale de l'équipe il y a un travail de gestion supplémentaire (assuré par M. Krafft)



Concernant la formation il est prévue un sous-traitance à partir de 1992; ceci entrainera aussi un travail de gestion supplémentaire et il faudra probablement prévoir un poste B supplémentaire pour l'équipe 3. FORMATION; cette même observation s'applique aussi pour la diffusion en général qui nécessite des ressources supplémentaires pour la coordination avec l'OPOCE et EUROBASES.

Concernant l'encodage externe la situation est comme suite:

l'on dispose 712,7 kECU;

la demande de 2,3 MECU pour l'extension de la couverture et rattrapage n'as pas été honorée.

Le fait que la gestion informatique locale et centrale est assurée par des équipes séparées (et pour la gestion centrale géographiquement éloignée) entraine un travail supplémentaire de gestion assuré en surcharge par Mme Hubert.

#### Modernisation

Le suivi du Projet de Modernisation nécessite les ressources suivantes:

Achevement TEXTERFACE, lancement 2ème sous-projet (ARCHIVE / BDDs): nécessite 1 fonctionnaire A Informaticien/Manager. La fonction est assurée par M. Alevantis en collaboration avec Mme Bensch (en parallèle avec d'autres tâches);

Suivi des sous-projets UFI et Didacticiel: nécessite 1 fonctionnaire A Informaticien/Manager. La fonction est assurée par M. Dewez (en parallèle avec d'autres tâches);

Le développement est assuré par l'équipe de M. Papacostas (ISI-BDI) comme est aussi le cas pour la maintenance évolutive et corrective; l'éloignement géographique entraine la consommation des ressources supplémentaires pour le suivi des actions. Les ressources modernisation et maintenance figurent en Annexe.

#### TOTAUX

Besoins:	40 personnes/années
	1 personne/année LSA
	2 personnes/années gestion centrale
	7 personnes/années sous-traitance analyse
	9 personnes/années sous-traitance formation
	2,3 MECU encodage externe
Disponibles	13 personnes/années (dont 1 - Mme Beviere - en prêt)
	1 personne/année LSA
	2 personnes/années gestion centrale CC
	5 personnes/années sous-traitance analyse
	6 personnes/années formation-documentation
	712,7 kECU encodage externe

C O N C L U S I O N

L'équipe CELEX doit être renforcée avec:

1 A 6/7	Manager Production
1 A 6/7	Juriste Analyste
1 A 6/7	Juriste Répertoire
2 A 6/7	Managers Modernisation
9 A/LA/B	Qualité des données
9 A/LA/B	Codification - Documentation

(la formation étant assurée dans le cadre du contrat Sdt et ensuite en sous-traitance, .... ECU)

2 B	Gestion locale
1 B	Serveurs

(si l'on reprend la gestion centrale il faudra 2 B/C supplémentaires)

1 C	Secretariat
-----	-------------

2 x 40 kECU pour 2 analystes supplémentaires

(il est entendu que les personnes chargées de la qualité des textes, de la documentation et de la codification, effectueront leur travaux de dactylographie eux mêmes; sinon il faudra renforcer l'équipe d'encodage de quelques 5 personnes C).

TOTAL : (sans reprise gestion centrale ni LSA)

5 A
18 A/LA/B
3 B
1 C
80 kECU analystes
.. kECU formation

Il faudra aussi regulariser les postes suivantes:

1 A 6/7	Juriste Codification	p. Devos (conversion en A)
1 A 6/7	Manager Formation	p. Krafft
1 C	Dispatching	p. Beviere

P. Alevantis

ANNEXE

Ressources Modernisation-Maintenance CELEX

1994		1991	1992	1993
<u>Unité ISI-BD</u>				
Maintenance				
ancien système (p.m.) kECU	70	50	50	
Texterface (achevement)	50			
Texterface (maintenance)		30	30	50
ARCHIVE/BDDs CO	150	250	100	
ARCHIVE/BDDs RI/maintenance			50	50
<u>Unité BD-CELEX</u>				
Renfort analyse	40	40	55	
Renfort gestion		45		