

IX.f.4 - S.I.I.
Bases de données

Bruxelles, le 24 novembre 1988

PROJET DE PREANALYSE
POUR LE DEVELOPPEMENT ET LA MODERNISATION
DU SYSTEME CELEX

novembre 1988



INTRODUCTION

I. SITUATION ACTUELLE

II. OBJECTIFS

III. SPECIFICATIONS DU NOUVEAU SYSTEME

A. LES BASES DE DIFFUSION

- A.1 Actualité et couverture
- A.2 Structure
- A.3 Multilinguisme
- A.4 Tableaux et graphiques
- A.5 Logiciel de diffusion
- A.6 Ressources à prévoir pour l'extension de la couverture

B. LA BASE DE PRODUCTION

- B.1 La nouvelle archive - Alimentation des données analytiques
- B.2 La nouvelle "interface" - Transfert des données analytiques
- B.3 Le "texterface" - Alimentation des textes

C. LA STRUCTURE DE L'APPLICATION ET SES RELATIONS AVEC IDS

- C.1 Solution proposée
 - A. Architecture
 - B. Description fonctionnelle
 - 1. Caractéristiques des bases
 - 2. Relations entre les bases
 - 3. "Texterface" et base de production des textes (BDPT)
 - 4. Alimentation et gestion de la base de production (archive)
 - 5. Sélection et transfert des documents
 - 6. Modification du texte d'un document
 - C. Configuration HW/SW
 - 1. Système local
 - 2. Système central
 - 3. Observations
- C.2 Relation avec IDS
- C.3 Solutions alternatives

IV. LE NOUVEAU SYSTEME EN FONCTIONNEMENT COURANT

A. L'ACCES EN LIGNE

/u/hubex/mod/preana

B. L'UTILISATION EN DIFFERE

C. LA MISE A JOUR

D. STATISTIQUES

- D.1 Statistiques de production
- D.2 Statistiques d'utilisation

E. DOCUMENTATION

V. DEVELOPPEMENT ET MISE EN PLACE DU NOUVEAU SYSTEME

A. STRATEGIE DE DEVELOPPEMENT

B. STRATEGIE D'IMPLANTATION

C. ESTIMATION DES RESSOURCES

VI. ETUDE DE FAISABILITE

VII. COOTS ET BENEFICES

ANNEXES

INTRODUCTION

Le système CELEX (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.

C'est la Commission qui 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.

D'autre part, dans sa résolution du 26 novembre 1975⁴, relative à l'automatisation de la documentation juridique (JOCE C 20 du 28 janvier 1975), le Conseil a décidé la création d'un comité d'experts composé de représentants des Etats membres et des Institutions européennes. Connu sous la dénomination de "Groupe Informatique juridique du Conseil", ce comité a reçu, entre autres, le mandat de suivre les développements de système CELEX, dont l'exécution a été confiée à la Commission en collaboration avec les autres Institutions.

Le système couvre l'ensemble du droit communautaire, réparti en 10 "secteurs" précisément définis (deux de ces secteurs sont encore à créer).

La base de données a pour vocation de permettre 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".

Le développement original de l'application s'est effectué sur technologie IBM (logiciel STAIRS). En 1976, sur décision de la Commission, une première conversion de CELEX a été entamée sur une machine ICL. Différents obstacles auxquels s'est trouvé confronté ce projet ont conduit, en 1980, à la signature avec BULL d'un contrat confiant la gestion des bases de données documentaires au service-bureau EURIS (ordinateur BULL IRIS 80 puis DPS 7).

En 1987, après la décision de reprise des applications EURIS par le Centre de Calcul de Luxembourg, CELEX a fait l'objet d'un transfert sur une machine BULL DPS 88, puis en 1988 sur un DPS 90 utilisant le logiciel documentaire MISTRAL V5 sur système d'exploitation GCOS 8.

Ouvert successivement aux Institutions puis au public, CELEX est maintenant utilisé depuis plus d'une dizaine d'années. La forme sous laquelle le système se présente actuellement est le résultat d'une longue évolution.

Les structures de l'application d'origine, sous format IBM, se retrouvent encore dans le noyau de la base de production des données ("l'archive"), et la plupart des développements réalisés depuis ont dû tenir compte de cet état des choses. 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.

Cette situation, ainsi que la dernière opération de conversion, ont constitué les principaux obstacles à la menée à bien du dernier plan triennal de développement (1983 - 1985) qui avait pour objectifs majeurs :

- l'achèvement du multilinguisme;
- le chargement de l'ensemble des textes intégraux dans la base;
- le développement de nouveaux secteurs

(voir à ce sujet le "Rapport de la Commission sur les développements du système CELEX", document 6496/86 soumis au Groupe Informatique juridique du Conseil le 13 mai 1986).

Il est maintenant 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.

La structure de la nouvelle application devra être modulaire, afin de rendre possible l'adaptation en souplesse aux progrès technologiques.

Les buts à atteindre consistent particulièrement en :

- convivialité de l'accès à l'information juridique;
- couverture plus cohérente et meilleure actualité;
- souplesse de l'administration du système;
- rationalisation des procédures autorisant la réduction des coûts d'exploitation;
- qualité de service accrue pour les différentes catégories d'utilisateurs et d'intervenants.

La réalisation de ces objectifs permettra de fournir un système d'information et de documentation juridiques de référence aux professionnels du droit ainsi qu'aux responsables économiques et administratifs dans les Etats membres. Il appartiendrait ainsi aux Institutions - et notamment à la Commission - d'apporter un élément supplémentaire à la meilleure implication du public européen dans la mise en place du Grand Marché.

Par ailleurs, des bénéfices directs et concrets en découleront pour les services de la Commission et des autres Institutions. Au-delà de la fourniture à moindre coût d'un service amélioré, les développements envisagés permettront d'assurer une productivité accrue, à la fois des différentes équipes intervenant dans le fonctionnement de CELEX et des nombreux utilisateurs internes de l'application.

I. SITUATION ACTUELLE

1. La complexité de CELEX est un phénomène depuis longtemps connu et reconnu. Le problème a, au moins une fois, fait l'objet d'un exposé concret et détaillé dans l'étude effectuée pour le compte de la Commission par la firme BUTLER COX en 1985 (General access to legal data bases or systems in Europe - Final study report - 05.07.1985). Bien que l'échantillonnage statistique utilisé pour CELEX puisse être discuté, les conclusions de l'étude ont été largement diffusées et ont rencontré l'adhésion de la grande majorité des utilisateurs de CELEX.
2. Tel qu'il se présente à l'utilisateur, les principales caractéristiques de CELEX peuvent être résumées comme suit :
 - a) 109 190 unités documentaires (au 27 juillet 1988), réparties sur 8 des 10 secteurs prévus. La répartition par secteur étant la suivante :

- secteur 1 (Traités)	: 2 367
- secteur 2 (Accords)	: 1 184
- secteur 3 (Droit dérivé)	: 36 441
- secteur 4 (Droit complémentaire)	: 407
- secteur 5 (Travaux préparatoires)	: 10 811
- secteur 6 (Jurisprudence)	: 4 830
- secteur 7 (Mesures nationales d'exécution)	: 12 860
- secteur 9 (Questions parlementaires)	: 40 290
 - b) 31 lexiques, répartis sur 80 champs (voir annexe 1 A).
 - c) Mise à jour hebdomadaire (voir statistiques en annexe 1B).
 - d) Disponible en français, anglais, allemand, néerlandais et italien, les versions danoise et grecque étant en préparation. Les versions linguistiques autre que française font l'objet d'une "couverture différenciée", certains documents donnant lieu à un filtrage de données avant chargement (voir annexe 1E).

Les versions disponibles peuvent être interrogées en mode TTY via une grande variété de terminaux (voir annexe 1C), mais le jeu de caractères pris en compte est le jeu de base 7 bits ASCII, c'est à dire sans les caractères "nationaux" (accents, etc ...). Avant chargement dans les bases, tous les textes subissent un traitement d'appauvrissement typographique. Il n'est pas, à l'heure actuelle, possible de reproduire à l'écran certains tableaux complexes ni aucun schéma pouvant être inclus dans le texte.
3. Un diagramme représentant les différents composants de système tels qu'il se présentent pour l'équipe responsable de l'administration de l'application figure en annexe 1D. L'introduction des données s'effectue par les chaînes de traitement et d'alimentation suivantes :
 - 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.

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

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

112 programmes
83 procédures
74 JCL's
116 fichiers (dont 68 utilisés pour la diffusion aux serveurs externes).

4. En termes d'utilisation, les différentes versions de CELEX totalisent le taux d'interrogation le plus élevé des bases documentaires de la Commission. Une ventilation des interrogations de CELEX par les utilisateurs internes et externes aux Institutions figure en annexe 1G.

En plus de l'interrogation en ligne, CELEX est utilisé pour :

- la production de "profils utilisateurs" (une vingtaine par semaine);
- l'extraction de données pour des sites serveurs externes;
- l'édition du "Répertoire de la législation communautaire en vigueur";
- la production du "système d'alerte en vue d'éviter l'apparition de vide juridique";
- l'alimentation de la base ASMODEE (contrôle de l'application des directives par les Etats membres).

II. OBJECTIFS

Les principaux objectifs du plan actuel de développement et de modernisation du système CELEX ont été exposés dans le "problem statement" présenté au Groupe Informatique juridique du Conseil lors de sa réunion des 11 et 12 juin 1987. Ils consistent en :

- la menée à terme des objectifs du plan précédent (1983 - 1985) : multilinguisme, développement des secteurs manquants, parachèvement de la couverture des textes intégraux pour la législation et la jurisprudence;
- l'amélioration de la convivialité d'accès à l'information, tant en ce qui concerne la couverture et l'actualité de celle-ci qu'en ce qui concerne la structure et la présentation des données analytiques et textuelles.
- l'amélioration de l'alimentation, de la mise à jour et de la gestion générale du système.

1. L'achèvement du plan précédent comprend notamment :

- la mise en place de la version danoise (prévue pour septembre 1988);
- la mise en place de la version grecque (prévue pour courant 1989);
- la création des versions espagnole et portugaise (en fonction de la disponibilité de ressources);
- la création du secteur 8 (jurisprudence nationale se référant au droit communautaire);
- la création du secteur 10 (doctrine publiée relative au droit communautaire);

2. L'amélioration de la convivialité d'accès à l'information doit se fonder sur :

- l'enrichissement du contenu de la base, aussi bien en ce qui concerne les données que leur degré d'actualité;
- la réduction du temps moyen nécessaire à un utilisateur pour accéder à l'information recherchée, réduction obtenue à travers une meilleure présentation des données et une normalisation de la "terminologie" employée dans la base elle-même (noms de champs standardisés, utilisation d'un thésaurus commun, mise en oeuvre étendue de la relation de synonymie, etc ...). Par ailleurs, il va sans dire que la mise en place d'une interface conviviale d'interrogation (UFI) constituera l'une des clés essentielles d'un meilleur accès à l'information.

3. L'optimisation des processus d'alimentation, de mise à jour et de gestion aura un impact immédiat sur la pertinence, la fiabilité et l'actualité des données.

Les objectifs à réaliser dans ce domaine sont :

- la mise à disposition d'un instrument d'analyse en ligne, avec contrôle formel des données introduites, la rationalisation et la simplification des traitements, et enfin, la normalisation des formats de transfert des documents par l'utilisation des standards ISO 2709 ou SGML;
- la rationalisation des procédures de mise à jour et de gestion, autorisant une réduction sensible des coûts d'exploitation et de maintenance.

Le développement substantiel du système CELEX a été arrêté en 1985 afin de limiter les difficultés et les risques de l'opération de transfert vers le Centre de Calcul de Luxembourg. De ce fait, il n'a pas été possible à la Commission de tenir son engagement vis à vis du Groupe Informatique juridique du Conseil de lui soumettre un nouveau plan triennal de développement pour la période 1986 - 1988. Il n'en reste pas moins avéré que l'évolution du système CELEX ne peut être envisagée dans les mêmes conditions que par le passé : les opérations de maintenance évolutive ou même corrective sont devenues, du fait de la complexité et de la fragilité de l'édifice, trop coûteuses - comme, d'autre part l'exploitation et la gestion courante de l'application dont la taille et la "lourdeur" se sont accrues sans que les adaptations technologiques nécessaires aient pu être effectuées.

Tout retard apporté à la modernisation de CELEX se traduira par :

- l'inadaptation croissante du système aux structures de l'architecture informatique de la Commission;
- l'impossibilité de profiter des solutions techniques développées dans le cadre du projet IDS/UAS, comme par ailleurs celle de contribuer à la mise en oeuvre effective de ces solutions,
- une perte de crédibilité de la Commission vis à vis du Groupe Informatique juridique du Conseil et des autres Institutions;
- une perception négative de la part des milieux professionnels concernés et du public européen en général, dont les besoins tels qu'exprimés dans l'étude BUTLER COX de 1985 ne sont toujours pas pris en compte;
- l'emploi injustifié de ressources à la maintenance et à l'exploitation d'un système dépassé.

d'ici

III. LES SPECIFICATIONS DU NOUVEAU SYSTEME CELEX

Le présent travail de préanalyse suivra la démarche décrite ci-après de façon à permettre 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 nécessaires à la mise au point 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 clientes", 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 suivi de l'exploitation courante dans les meilleures conditions d'économie et de transparence ("manager friendliness").

Pour chacun de ces ensembles de spécifications, il sera tenu compte de la possibilité éventuelle d'intégrer les développements correspondants dans le contexte du projet IDS. Il convient par ailleurs de noter, à ce stade de la préanalyse, que certaines des annexes jointes au présent document constituent une part importante de définition des problèmes et des solutions qui pourraient être envisagées; il va de soi qu'il s'agit de documents de travail destinés à être complétés et finalisés au cours de l'étude de faisabilité.

Monsieur Hubert, Vanhove, Benière

pour les spécifications de la nouvelle base de diffusion

pu/falero/mob/spec

la partie sur la base de diffusion

Monsieur

Hubert

Vanhove

A. LES BASES DE DIFFUSION

A.1 Actualité et couverture

L'agrément d'utilisation de la base de données peut d'abord être accru par l'enrichissement de son contenu au moyen d'une meilleure actualité et d'une couverture plus complète.

L'amélioration de l'actualité implique l'accélération des mises à jour des différents secteurs de CELEX. Les données analytiques (mais aussi les textes) devraient être disponibles à l'écran dans la semaine suivant la publication de l'acte. En ce qui concerne la couverture, il faut travailler à compléter et harmoniser les différents secteurs et les différentes versions linguistiques eu égard à :

- l'existence des documents dans les différentes versions linguistiques;
- la couverture dans le temps des différents secteurs;
- la disponibilité de traductions officielles des textes intégraux.

Analysée par secteur, la couverture du nouveau CELEX devrait être la suivante (les estimations du volume des textes intégraux devant être ajoutés figurent en annexe 2A) :

- 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 champs spécifiques destinés à accueillir les références des mesures nationales d'exécution de ces directives (en remplacement de l'actuel secteur 7);
- 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 (ce texte devant être effacé après l'adoption de la proposition et son apparition en tant qu'acte du secteur 3);
 - introduction éventuelle du texte des avis du P.E.;
- Secteur 6 (Jurisprudence) :
 - extension éventuelle de la couverture aux saisines, et affaires radiées;
 - 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 (actuellement : Mesures nationales d'exécution) :
 - suppression du secteur 7 actuel, et transfert de l'information relative aux MNE dans des champs appropriés rattachés aux directives du secteur 3.
 - étude de la possibilité d'utiliser le secteur 7 devenu "vacant" pour rendre accessible aux utilisateurs une version des actes juridiques intégrant dans le texte leurs modifications successives (à étudier, le cas échéant, dans une phase ultérieure du projet de modernisation).
- 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 (Question 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.

La systématisation de la "couverture différenciée" doit disparaître. Il sera établi que chaque document figurera dans les bases avec son texte, dans la mesure où celui-ci aura été traduit et sera disponible sur support magnétique. Un effort de rattrapage sera consenti pour les textes non-disponibles, lors de leur création, sur de pareil supports.

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, et avec une mention appropriée pour l'information des utilisateurs.

A.2 Structure

La complexité de la structure actuelle de CELEX se traduit d'abord pour ses utilisateurs par une difficulté d'assimilation et de mémorisation de tous les paramètres. Une première étape dans la voie de la rationalisation consisterait à adapter ces paramètres (noms des lexiques, des champs, etc ...) aux normes et règles le plus généralement admises ou pour lesquelles un processus de standardisation est en cours dans le domaine des bases de données documentaires. D'autre part, il existe une relation directe entre la complexité de la structure, telle qu'elle est perçue par les utilisateurs, et la disposition des données dans la base.

La performance de l'accès à l'information est tributaire de la présentation de cette dernière, dont l'amélioration doit tenir compte de l'expérience acquise dans l'interrogation des bases documentaires, des besoins exprimés par les utilisateurs et des nouvelles possibilités de la technique. Les données formatées devront être à la fois plus claires et plus cohérentes alors que les textes devront offrir à la fois une meilleure lisibilité et un meilleur taux de précision et de rappel lors de l'interrogation.

De fait, de par sa vocation de base de données de type "texte intégral", on pourrait concevoir que CELEX ne comporte qu'un nombre réduit de rubriques analytiques. Toute l'information juridique relative à un acte est en principe présentée dans le corps de celui-ci tel qu'il est publié, et devrait donc pouvoir être recherchée et extraite par interrogation sur le texte intégral du document. Cependant, cette information n'est pour ainsi dire pas structurée, et ne le sera pas à terme prévisible.

Son utilisation comme critère de recherche nécessiterait donc la mise en oeuvre d'un système expert sophistiqué faisant appel à des techniques d'intelligence artificielle qui restent à développer. De plus, outre les risques de "silence" en cas d'absence d'un texte, certains éléments d'information tels que la classification par mots-clés, le nom du juge rapporteur, la date explicite de fin de validité, etc ... n'y figurent pas et leur introduction dans les rubriques bibliographiques est le résultat du travail d'analyse effectué au sein des différentes équipes chargées de l'alimentation de CELEX dans les Institutions.

Pour offrir le maximum de possibilités de recherche, la base de données a été conçue dès l'origine pour contenir une partie textuelle et une partie analytique, cette dernière devant faciliter l'accès à l'information via l'emploi d'un vocabulaire contrôlé. La redondance d'information entre les deux parties est dès lors inévitable, mais volontaire. Ce qui est par contre inacceptable est que la partie analytique comporte une part non négligeable de redondance sur elle même. Ce problème doit être résolu dans la nouvelle version de CELEX, dans laquelle les champs analytiques seront regroupés en six fichiers inverses dont la description détaillée fait l'objet de l'annexe 2C.

A.3 Multilinguisme

Comme il est rappelé en introduction, CELEX doit être disponible dans chacune des neuf langues officielles des Communautés. La mise en place du multilinguisme intégral pose plusieurs problèmes. Un utilisateur de la base française devrait disposer à l'écran des textes dans leur forme "normale", c'est à dire avec toutes les spécificités typographiques de la langue. Ceci s'applique aussi aux versions allemande, danoise et espagnole, sans parler, bien sûr de la version grecque.

D'autre part, comme 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, il est impératif de préserver toutes possibilités de reconnexion, sauvegarde, etc ...

La prise en compte de ces considérations implique qu'une attention particulière soit portée aux tables de caractères qui seront utilisées dans les bases, depuis l'encodage des données jusqu'à leur interrogation en-ligne. L'architecture informatique des Institutions prévoyant l'utilisation de terminaux au standard VT220, il sera possible de bénéficier des avantages offerts par ce type de matériel.

La question du multilinguisme doit être examinée sous deux aspects :

- quelle codification adopter pour les données qui alimentent les bases CELEX;
- comment rendre ces données à l'écran avec leur richesse typographique (sans "appauvrissement").

En ce qui concerne la codification, le principe de travail repose sur la présence dans les bases interrogeables de toute la richesse typographique attachée aux données. Déjà, dans sa version la plus récente, le logiciel de recherche documentaire utilisé pour les bases de diffusion CELEX fonctionne en "mode 8 bits" ce qui permettrait dès à présent de codifier l'information en utilisant les normes ISO 8859/1 pour les caractères "latins" et ISO 8859/7 pour les caractères grecs. Cette solution ne permettra sans doute 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 des 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 de recherche. Le cas échéant, rien n'empêcherait d'opérer une translittération "intelligente" de ces caractères, comme c'est le cas actuellement pour les symboles mathématiques.

A.4 Tableaux et graphiques

La restitution sur un écran 80 x 24 ou 132 x 24 des tableaux et graphiques pouvant figurer dans un texte publié 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 est nécessaire dans une première étape, la seconde relevant des possibilités du logiciel documentaire de traiter et mettre en forme les données. A l'heure actuelle, cette mise en forme est réalisée - dans la mesure du possible et pour les tableaux de taille réduite - lors de l'encodage.

A.5 Logiciel de diffusion

Etant donnée la situation actuelle de CELEX et les caractères spécifiques de l'application, le choix du logiciel pour les bases de diffusion revêt une certaine importance. A ce stade de l'évolution du projet, il ne paraît cependant pas nécessaire de fixer une option a priori. Dans sa conception d'ensemble, le nouveau CELEX marquera clairement la distinction entre la diffusion des données et leur production, et cette dernière s'organisera de manière à préserver l'indépendance de l'application par rapport aux logiciels documentaires.

Le moment venu, le choix définitif devra s'effectuer en tenant compte aussi bien des fonctionnalités requises que de leur disponibilité opérationnelle. Il conviendra aussi, sur base de la situation existante en matière d'utilisation finale, de procéder à une analyse des coûts et bénéfices selon la solution envisagée. Ce travail devra s'insérer dans l'étude de faisabilité.

A.6 Ressources à prévoir pour l'extension de la couverture

L'introduction dans CELEX de données complémentaires, telle que décrite ci-dessus, implique l'allocation de différents types de ressources :

- a) Le temps supplémentaire nécessaire à l'analyse, ainsi que le volume prévisible des données sont exposés dans l'annexe 2B. Le rattrapage des textes disponibles sur support magnétique est en cours, l'Office s'étant engagé dans ce sens à la suite d'une demande déjà ancienne formulée par CELEX.

- b) De nouvelles méthodes d'acquisition des données devront être mises en place dans le futur. Les modalités techniques sont exposées dans le chapitre relatif à la nouvelle interface "textes". La planification des ressources pourra être établie en fonction des solutions retenues.

B. LA BASE DE PRODUCTION

Bien que n'étant pas directement perçues par les utilisateurs, l'alimentation et l'administration des bases de diffusion et de production mérite une attention particulière car leur efficacité conditionne la qualité, la fiabilité et l'actualité de l'application.

L'archive CELEX aussi bien que les différentes chaînes d'alimentation et de traitement des données doivent être améliorées et souvent entièrement repensées.

Les grands composants des chaînes d'alimentation et de traitement (introduction des données analytiques et des données textuelles, interfaces entre bases de production et de diffusion, etc ...) sont à reconcevoir pour aboutir à un système modulaire et facile à maintenir, dont les caractéristiques principales seront :

- un meilleur contrôle des données introduites;
- l'analyse et la correction en-ligne des documents;
- des statistiques détaillées et complètes;
- de meilleures possibilités d'accès à l'archive (compte tenu du caractère interinstitutionnel de CELEX);
- des traitements simplifiés;
- la normalisation aussi poussée que possible des formats de transferts des documents (ISO 2709, SGML, ...).

Tous ces éléments constituent la base de la rationalisation de l'administration du système.

B.1 La nouvelle "archive" - Alimentation des données analytiques

Les procédures actuelles d'introduction des données dans CELEX sont compliquées, coûteuses en ressources, de maintenance évolutive ou corrective difficile.

Néanmoins, le concept même de "l'archive neutre" correspond parfaitement aux nouveaux principes de la Commission en matière de développement de systèmes d'information dans la mesure où ce fichier constitue une véritable base de production des données dont les version linguistiques de CELEX sont les relais de diffusion.

Les lignes directrices du développement du nouveau système d'alimentation des données analytiques doivent être :

- Maintien de l'archive dans son rôle de base de production, accessibles par les membres des équipes CELEX au sein des Institutions (plus ou moins 25 personnes). Les avantages offerts par une archive centrale ne sont plus à démontrer (production de sept - et bientôt neuf - versions linguistiques à partir d'un noyau unique, contrôle de la cohérence plus facile, réduction du risque de captivité de l'application, etc ...) et le principe lui-même ne sera pas remis en question. La nouvelle version de l'archive accueillera les données destinées à l'alimentation des champs analytiques décrits en annexe 2C.

- Mise en place d'un système d'analyse documentaire en-ligne. Il ne devrait pas y avoir de difficultés à définir des écrans de saisie correspondant aux actuels bordereaux d'analyse et disposant, selon les zones, de modules de contrôle en table des informations introduites. L'annexe 2E donne une liste indicative des bordereaux à traduire sur écran. Par ailleurs, certains contrôles plus poussés devraient pouvoir être mis en oeuvre (p.ex. contrôle de la cohérence des données entre deux zones, etc ...). L'annexe 2D fournit plus de détail sur les procédures de contrôle et de traitement des rubriques analytiques de l'archive. Il convient aussi de prévoir des procédures d'alimentation de l'archive en traitement différé (passerelles entre application, p.ex. d'ASMODEE vers CELEX, etc ...) à la condition que les données candidates soient disponibles en format normalisé (ISO 2709). Un soin particulier doit être accordé à la procédure de suppression des données qui, à l'heure actuelle, requiert une chaîne de traitement spécifique (chaîne ANNULATIONS) et une gestion très peu pratique. Le retrait d'un document de l'archive doit entraîner automatiquement, outre sa disparition des bases, l'annulation des citations dont il fait l'objet dans d'autres actes.
- Facilité d'ajout de nouveaux champs dans la base. Cette requête peut sembler contradictoire avec l'idée généralement admise qu'une base documentaire doit avoir une structure immuable. Ce serait ignorer la dimension juridique d'un système comme CELEX qui doit, par vocation, s'adapter à l'évolution du droit communautaire. Dans le passé, et encore très récemment, cette exigence a imposé des efforts coûteux pour apporter les modifications nécessaires. Le nouveau système devra tenir compte de cet impératif et offrir une souplesse fondée sur des possibilités maximales de paramétrage et une construction modulaire, de sorte qu'une modification n'entraîne plus de gros travaux de programmation (un bon exemple pratique est à tirer de la nouvelle version de la chaîne INFOCELEX).

D'une façon générale, il doit être possible d'opérer sur l'archive les manipulations nécessaires (ajouts, retraits et modifications ...) et de retracer ces modifications de manière à effectuer les mises à jour nécessaires des bases de diffusion (rôle actuellement joué par le fichier NEWDATA). Les outils statistiques et comptables nécessaires font l'objet d'une description séparée. Il consistent essentiellement, pour cette partie, en statistiques détaillées sur les nouveaux documents introduits, les retraits, les modifications dans les différents champs, etc ...

B.2 La nouvelle "interface" - Transfert des données analytiques vers les bases de diffusion

La nouvelle interface à développer devra offrir les fonctionnalités suivantes :

- transfert des données à partir de l'archive vers les fichiers appropriés du logiciel de recherche documentaire utilisé pour les bases de diffusion. En amont des bases de diffusion, les fichiers de mise à jour doivent être au format ISO 2709 pour minimiser le risque de captivité des chaînes d'alimentation par rapport au logiciel documentaire.

- module de gestion et mise à jour de thésaurus. Cette requête se justifie, outre les besoins particuliers découlant du multilinguisme, par le fait que la nouvelle structure de la base nécessitera beaucoup plus de manipulation de thésaurus qu'à l'heure actuelle; la réduction des redondances d'information implique la mise en oeuvre "en grand style" des relations de synonymie, de hiérarchie et de groupe entre les différents éléments d'information.
- souplesse maximale dans la définition des paramètres de traitement. Si l'on prend, par exemple, une rubrique donnée dans l'archive, le responsable de CELEX doit disposer d'un système convivial (par menu) lui permettant notamment de définir le nom des champs correspondant dans les versions linguistiques ainsi que les types de traitement à effectuer lors du transfert entre la base de production et les bases de diffusion :
 - * absence de traduction (dates, texte libre, etc ...);
 - * traduction dans les langues par table de contrôle;
 - * mise en forme des données, par traduction de certains éléments, introduction de caractères spéciaux ou de libellés à certains endroits, transfert tel quel d'autres éléments selon leur position.

B.3 Le "texterface" - Alimentation des textes

Dans le contexte d'urgence générale du projet de modernisation de CELEX, la construction d'une nouvelle interface de chargement des textes doit revêtir une **priorité absolue**.

La première raison en est que, à court terme, tous les textes stockés jusqu'à présent sur bandes magnétiques devront être rechargés dans les bases. En leur temps, ces textes ont été chargés en format "appauvri" tandis que les bandes - qui contiennent l'information en "riche" - ont été conservées dans l'attente de disponibilité d'une version de MISTRAL pouvant prendre en compte ce genre de typographie. Cette fonctionnalité est désormais opérationnelle sur le DPS 90 de la Commission, et un plan de rechargement des textes est en cours d'élaboration afin de soulager le Centre de Calcul des contraintes de stockage d'un nombre toujours croissant de bandes dont, au fil du temps, il ne peut plus garantir la fiabilité des données.

D'autre part, l'Office des Publications est en train de mettre en oeuvre, dans les contrats de fourniture passés avec les imprimeries, le standard FORMEX dans le format duquel toutes les bandes issues de la photocomposition du Journal officiel devront lui être livrées. Le risque existe de devoir développer encore une chaîne d'alimentation des textes alors qu'il en existe déjà plusieurs. Il est donc rationnel et économique de prévoir - et de réaliser sans plus attendre - la refonte du système d'alimentation des textes. Celui-ci devra présenter les caractéristiques suivantes :

- possibilité de répartir les textes dans les champs appropriés, de façon automatique ou au moins semi-automatique;
- possibilité de téléchargement d'un texte vers une application locale permettant sa correction/modification et son retour vers la base centrale.

- possibilité de nouvelles équivalences pour certains caractères ou chaînes de caractères (transcodage et translitération). Il s'agit non seulement d'une conditions indispensable à la mise en oeuvre du multilinguisme tel que décrit ci-dessus, mais encore, de façon plus générale, d'un outil qui permettra d'adapter le système à l'évolution future des formats des données. L'expérience passée de l'application a montré que le format universel et définitif n'existe pas. Un système paramétrable permettrait de considérables gains de temps et d'argent.
- possibilité de prendre en compte des tableaux aux standards FORMEX ou SGML et, dans le futur, des éléments graphiques.
- possibilité de lien avec des modules d'analyse morphologique ou syntaxique (traitement linguistique) des textes au moment de leur indexation dans les bases. Des outils de ce type seront à rechercher sur le marché le moment venu. Il faut déjà prévoir des évolutions séparées en fonction de la langue traitée. Ce point ne constitue pas une priorité.
- enfin, possibilité de générer dans l'archive le bordereau d'analyse du document, en extrayant déjà le maximum de données analytiques possible. Bien qu'à l'heure actuelle la création de la partie analytique précède le chargement du texte, il faut considérer que celui existe sur support magnétique avant la parution du Journal officiel. Une fois créé, ce bordereau d'analyse électronique sera mis à disposition du DBM puis des analystes pour introduction des données complémentaires, contrôle et chargement dans la base de production pour s'insérer dans le processus normal des mises à jour des bases de diffusion. Après vérification du transfert, le chargement du texte proprement dit peut s'effectuer.

C. LA STRUCTURE DE L'APPLICATION ET SES RELATIONS AVEC IDS

C.1 Solution proposée

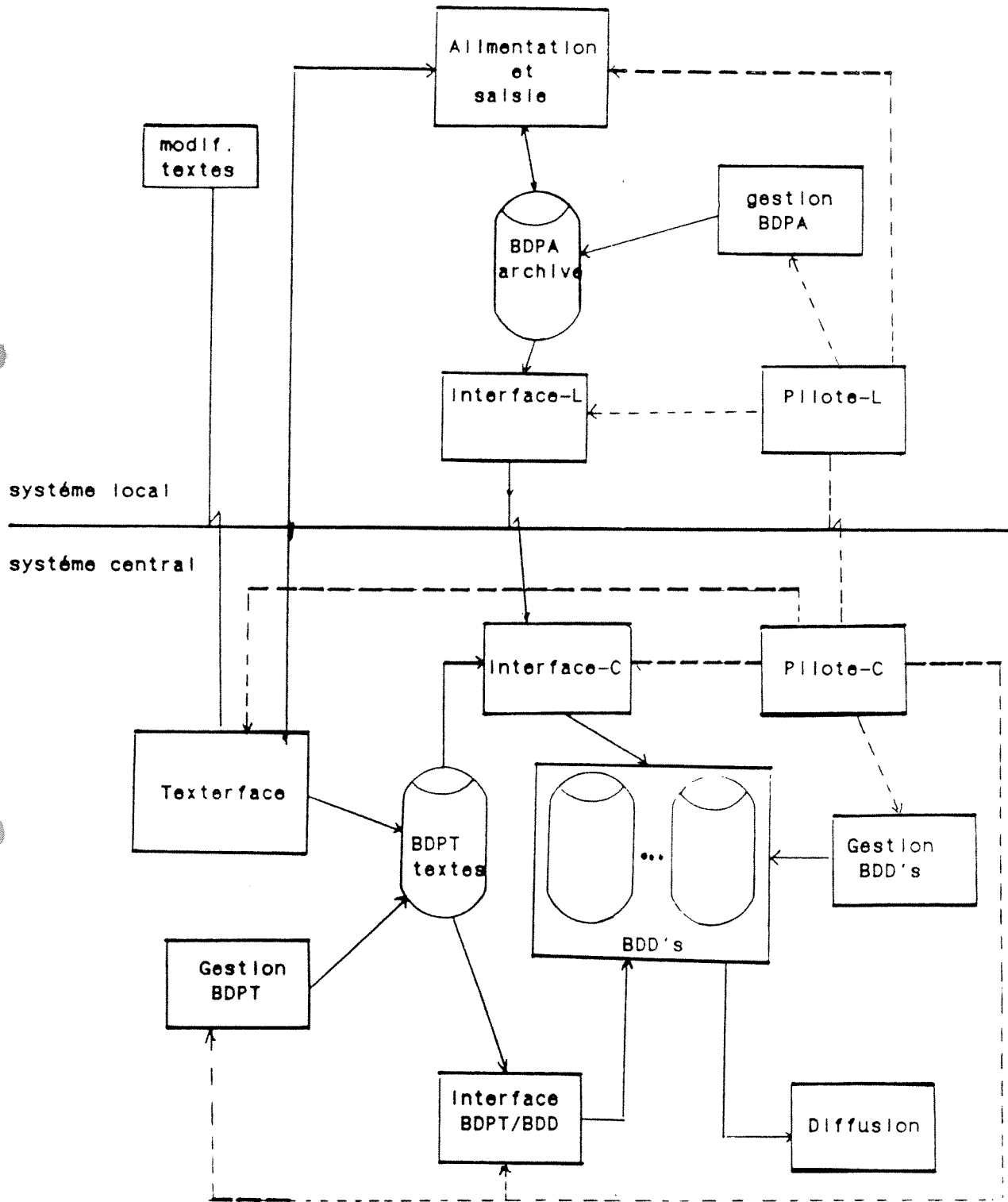
A. Architecture

D'après les spécifications ci-dessus, l'application CELEX peut se structurer selon les modules suivants :

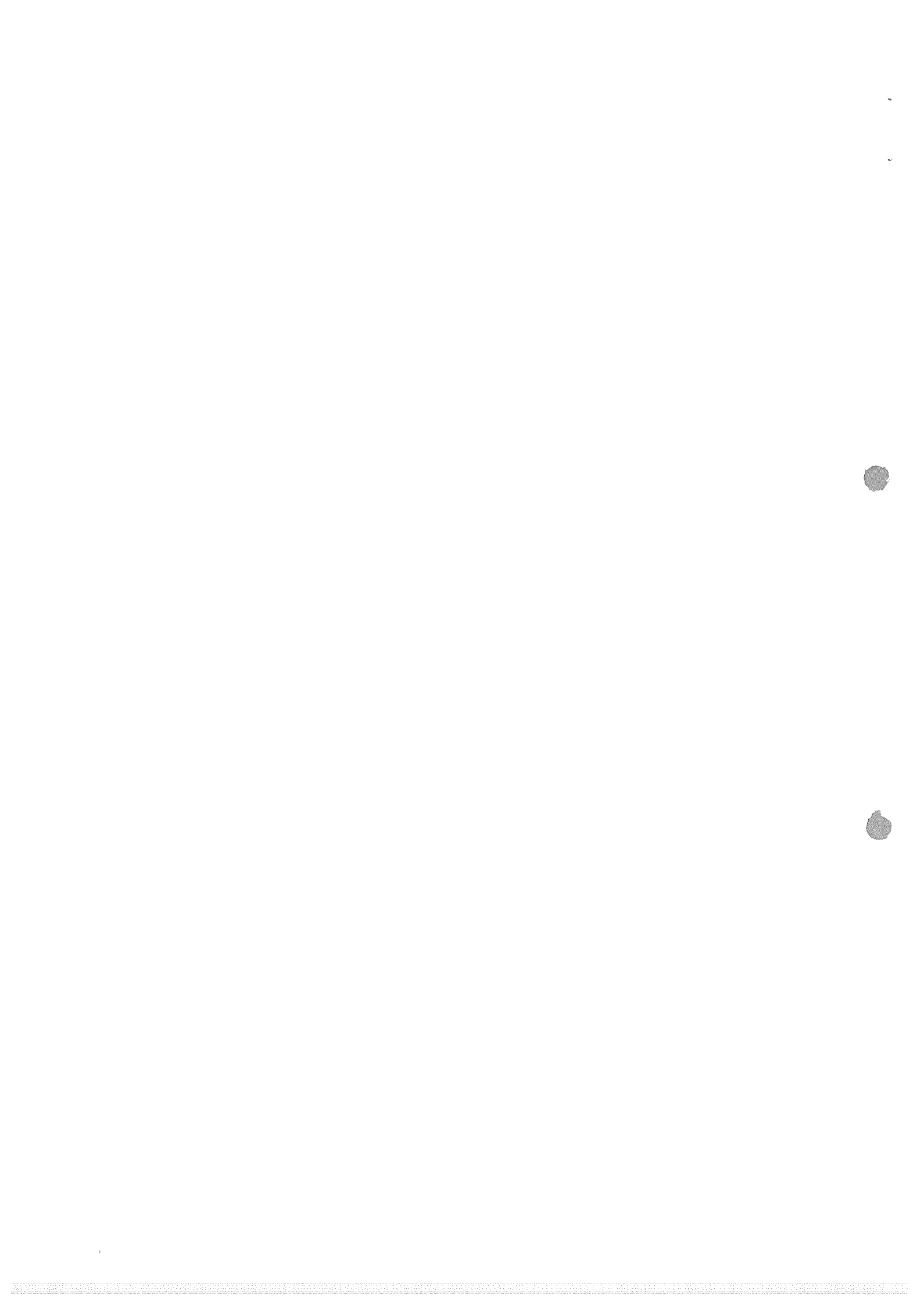
1. Alimentation/saisie des données analytiques : comprend toutes les fonctions (saisie en ligne, saisie complémentaire, passerelle avec d'autres bases, récupération de l'analyse des textes intégraux, ...) relatives à l'alimentation de la base de production (archive).
2. Gestion de la base de production (archive) : L'archive constitue le noyau fondamental pour la production des champs analytiques des différentes versions linguistiques CELEX. L'archive est indépendante des versions linguistiques et il existe une correspondance univoque entre les documents dans l'archive et dans les bases de diffusion. Les citations entre documents CELEX seront gérées à ce niveau et envoyées comme des modifications dans les bases de diffusion.
3. L'interface analytique : la fonction de ce module est de transformer un document "codifié" provenant de l'archive en un document CELEX pour chacune des versions linguistiques. Il doit également établir la liaison entre la partie analytique d'un document et la partie textuelle si elle est disponible.
4. Gestion des bases de diffusion : il y a une base CELEX différente pour chacune des versions linguistiques, et les tâches de gestion (m.a.j., sauvegarde, ...) doivent se faire en parallèle pour toutes les bases.
5. Diffusion des bases : le but de CELEX est d'être diffusé parmi un vaste ensemble d'utilisateurs. L'accès de tous ces utilisateurs doit être organisé, et facilité avec la mise en place des interfaces conviviales (UFI). Egalement, il faut prévoir la diffusion via supports magnétiques (bande magnétique, CD-ROM).
6. Gestion de la base de production des textes : les bases CELEX sont des bases documentaires à texte intégral. L'ensemble de textes en attente de chargement sera structuré autour d'une base de données temporaire (base de production de textes : BDPT) où les textes seront jusqu'à leur chargement définitif dans les bases CELEX.
7. Le "texterface" : la fonction de ce module est d'analyser les textes disponibles, construire les différents champs de type texte d'un document et si possible fournir le contenu de certains champs de type analytique qui seront répercutés au niveau du module de saisie pour être chargés dans l'archive.
8. Le "pilote" : l'application CELEX contiendra un ensemble d'éléments qui vont guider tant le processus de production (tables de contrôle et validation, tables de correspondance entre champs, tables de traduction, ...) et qui vont être gérés par le DBM. Le but de ce module est de mettre à disposition du DBM toutes ces fonctionnalités sous la forme d'une interface uniforme, cohérente et automatisée.

L'architecture générale de l'application CELEX à proposer est de type distribué, et elle est déterminée par la localisation et les relations entre les modules définis ci-dessus, les caractéristiques spécifiques de l'application et le cadre général prévu dans l'architecture informatique de la Commission. Ainsi :

- les bases de diffusion seront toujours sur machine centrale;
 - la base de production des textes et le module "texterface, étant donné le grand volume des textes à traiter et à transférer, seront sur le système central;
 - l'alimentation/saisie et la gestion de la base de production (archive) seront en local (c.à.d. sur une machine locale).
- Puisque l'archive (BDP) et les bases (BDD) seront sur des machines différentes, le module Interface aura une partie locale et une partie centrale. La partie locale produira un document, selon une structure propre à CELEX contenant les champs analytiques en forme "codifiée" et la partie centrale traduira ces codes dans les différentes versions linguistiques et fera le lien avec les parties textes disponibles de façon à produire un format standard (SGML), indépendant du logiciel de la BDD.
- Le pilote sera également organisé en deux parties, centrale et locale, bien que la présentation à l'utiliser devra être unique et avec un seul point d'entrée (en local).
- Il faudra un module de transfert d'information entre machine locale et machine centrale.
- Le schéma global de l'architecture proposé est le suivant :



Architecture CELEX (schéma général)



B. Description fonctionnelle

Il s'agit de décrire les fonctionnalités les plus importantes du système et la façon dont elle seront réalisées dans l'architecture proposée.

1. Caractéristiques des bases : dans l'architecture proposée, l'application est structurée autour de trois types de bases : la base de production ou archive (BDPA), les bases de diffusion (BDD's) et la base de production des textes (BDPT). Les caractéristiques de ces trois bases sont les suivantes :

BDD's :

- orientation : diffusion d'information
- accès public;
- base documentaire avec des documents à texte intégral;
- interrogation avec des fonctionnalités spécifiques de recherche documentaire;
- jeux de caractères "riche" (normes ISO 8859/1 et ISO 8859/7);
- possibilité de définir plusieurs champs longs (plus grands que 10 000 caractères) à longueur indéfinie;
- saisie en ligne : aucune.

BDPA (archive) :

- orientation : saisie et production de documents;
- accès : l'équipe CELEX et les Institutions responsables des secteurs CELEX;
- base structurée avec des champs textuels répétitifs (relation 1 → n) et aussi des champs de longueur indéfinie (entre 20 et 500 caractères);
- interrogation : par clé, avec un minimum d'accès par contenu;
- jeux de caractères riche optionnel
- saisie en ligne : fonctionnalité importante.

BDPT :

- base composée d'un certain nombre de champs de structure fixe plus un seul champ type "texte" de longueur illimitée;
- accès : par clé, avec visualisation du texte long;
- primitives d'accès aux fonctionnalités du logiciel depuis un langage de haut niveau;
- orientation : validation et vérification des textes stockés temporairement;
- saisie en ligne : seulement pour quelques temps. Les textes sont chargés par un processus batch;
- accès : l'équipe CELEX et les Institutions responsables des secteurs CELEX.

2. Relations entre les bases

Il faut définir les relations de données entre les différentes bases pour garantir la cohérence dans le fonctionnement de l'application.

- L'archive disposera des rubriques de gestion suivantes :

a) Relation avec les BDD's :

- . date de création du document
- . date de la dernière modification
- . flag de transfert
- . date du chargement/modification des champs analysés dans les BDD's

b) Relation avec la BDPT :

- . numéro provisoire avec la référence du texte dans la BDPT. Il y aura une référence par version linguistique
- . flag de récupération du texte

Les documents nouveaux ou modifiés dans l'archive seront transférés complets, tandis que pour les documents à supprimer on enverra seulement la référence. Tout mouvement (création, modification, élimination) sur l'archive sera stocké afin de pouvoir le répercuter correctement sur les BDD's.

- Dans la BDPT, on aura les champs de gestion suivants :

- . numéro de document CELEX (numdoc) modifiable après contrôle
- . numéro provisoire et code langue (clé) pour faire le lien avec la partie analytique de l'archive
- . date de chargement du texte dans la BDD
- . flag de transfert pour contrôle du processus de transfert entre BDPT et BDD
- . date de chargement/modification d'un enregistrement dans la base.

- Les bases de diffusion (BDD) disposeront des champs de gestion suivants :

- . date de création du document
- . date de m.a.j. des champs textuels
- . date de m.a.j. des champs analytiques

- Il serait convenable, pour garantir au maximum la cohérence entre les trois bases, que le résultat du processus de m.a.j. des bases de diffusion soit repercuté dans chacune des bases de production.

- Le transfert d'information entre la BDPT et les BDD's et entre l'archive (BDDA) et les BDD's se fera selon un format standard (ISO 2709, SGML).

3. "Texterface" et base de production des textes (BDPT)

Le processus de manipulation des textes des documents dans CELEX aura comme noyau une base documentaire (BDPT) sur le système central et dans laquelle les textes seront stockés temporairement avant leur chargement dans les BDD's.

Structure de la base

Chaque document aura au moins les champs suivants :

- numéro provisoire, alloué automatiquement pendant le processus de l'analyse du texte
- langue du texte
- numéro du document (num.doc) CELEX modifiable après contrôle
- flag de transfert

- date de chargement/modification dans cette base
- date de chargement/modification dans la base centrale
- titre du document
- texte; ce champ contiendra le texte complet du document mais structuré selon le format suivant :
 - <ch1>
 - contenu champ1
 - <ch2>
 - contenu champ2
 -
 - <chn>
 - contenu champ n

Cette base est unique et, en conséquence, valable pour toutes les versions linguistiques.

Alimentation de la base

Les bandes provenant des différentes sources (JO, Recueil) avec les textes à charger seront lues, traitées et analysées de façon à :

- construire le format interne à CELEX pour chaque version linguistique d'une publication. Chaque document sera identifié par un numéro provisoire (généralisé automatiquement) et le code de la langue.
- Pendant l'analyse de la version française d'une publication, un certain nombre de champs analytiques seront générés automatiquement parmi lesquels si possible le num.doc CELEX. Ces champs avec le numéro provisoire du document seront transférés vers le système local pour être injectés dans l'archive.
- Il sera également possible de charger l'introduction directe du num.doc dans cette base avec un mécanisme analogue à celui du fichier de concordance dans le CELEX actuel.
- La chaîne INFOCELEX pour construire les documents informatifs sur les BDD's pourra travailler sur cette base.

Gestion et interrogation

- Les informations disponibles dans la base seront accessibles par l'équipe de gestion CELEX via les fonctionnalités du logiciel documentaire choisi.
- La BDPT peut aussi avoir le rôle "d'historique" du processus de chargement des textes.
- Toutes les version linguistiques d'un document pourront être recherchées dans cette base, à l'aide, soit du numéro provisoire, soit du num.doc s'il existe.
- La gestion de cette base comporte certaines tâches parmi lesquelles on peut citer : le nettoyage des textes déjà chargés dans les BDD's, le transfert des résultats des analyses des textes vers le système centrale, etc ...
- Pour contrôler la procédure de nettoyage, la m.a.j. des textes dans les BDD's sera repercutée sur cette base.

4. Alimentation et gestion de la base de production (archive)

- La saisie des documents (partie analytique) se fera directement en ligne avec un ensemble d'écrans formatés selon le type de document à saisir. Ces écrans seront développés à l'aide du module de saisie du logiciel choisi pour implémenter la BDPA.
- 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 analyses, analyse automatique des textes) avec lesquelles le contenu de certains champs pourra s'obtenir automatiquement. Dans ce cas, les documents seront chargés provisoirement dans la base et, par après, on fera la saisie complémentaire et la validation de tout le document.
- L'analyse du texte d'une publication dans le module "Texterface" produira un fichier contenant certains attributs du document : numéro d'identification provisoire, num.doc CELEX, titre en français, ... Ce fichier sera transféré vers la machine centrale où il sera validé par l'équipe de gestion afin d'ajouter le num.doc, si nécessaire, et de faire la distribution entre les différents analystes qui vont compléter l'analyse du document.
- L'accès à l'information contenue dans l'archive pourra se faire :
 - a) avec le langage d'interrogation du SGBD choisi;
 - b) à l'aide de menus formatés.
- La structure du document ne sera pas exactement la même que celle qu'il y aura dans les bases de diffusion, cependant le DBM disposera d'une fonctionnalité pour lui fournir une présentation du document tel quel il est visualisé dans la BDD.
- Les tâches propres de gestion d'une base de données.
- L'équipe CELEX sera le responsable de la gestion.

5. Sélection et transfert des documents

a) Entre la BDPA et les BDD's

Les divers champs de type date et le champ transfert permettront d'établir les critères de sélection des documents à transférer. Les documents sélectionnés seront formatés selon une structure proche du document dans la BDD. Ce processus sera contrôlé par une table externe (qui définit la relation entre le format d'entrée et le format de sortie) et produira une forme disons "codifiée" du document qui sera envoyé vers le système central. Le format du document à ce niveau sera un format interne à CELEX.

b) Entre la BDPT et les BDD's

Le chargement des textes de la BDPT vers les BDD's peut se faire de la façon suivante :

* Sous le contrôle de l'interface

Le document codifié provenant de l'archive contiendra le numéro provisoire de chargement du texte dans le BDPT. Ce numéro, avec le code de langue, permettra de retrouver dans la base le texte associé. Deux possibilités de traitement :

- L'interface actualise la BDPT avec le num.doc (s'il n'existe pas) et marque le document comme transférable.
- L'interface récupère directement le texte pour construire les bordereaux de chargement dans les BDD's.

* Via un transfert direct entre les bases

Il peut exister une interface pour transférer les documents de la BDPT vers les BDD's. Cette interface sélectionnera les documents qui ont été marqués comme transférable (soit par l'utilisateur, soit automatiquement par le module interface) et construira les formats de chargements SGML pour les BDD's.

6. Modification du texte d'un document

- Il y aura toujours des modifications à faire sur la partie texte d'un document (texte proprement dit ou tout autre champ de nature textuelle), suite, soit à des erreurs de frappe, soit à des erreurs de gestion. Pour faire ces modifications, on aura un outil pour extraire le texte du système central, le transférer sur la machine locale, générer un document en format traitement de textes (Q1) et le renvoyer pour le recharger dans la BDPT. L'extraction du document se fera à l'endroit où a été repéré le document erroné, soit la BDD soit la BDPT.
- Le processus de création/modification d'un texte en local pour le rechargement postérieur dans la BDPT sera transparent pour l'archive.

C. Configuration HW/SW

1. Système local

Hardware

Une machine avec le système d'exploitation UNIX de l'ordre de 300 MB d'espace en disque et la possibilité de travail pour 20 utilisateurs dont 10 en simultané.

Software

- Le logiciel documentaire BASIS (release L ou K) avec ses différents modules (SCREEN, OLIVE, et les modules HLI (High Language Interface) pour avoir l'accès aux fonctionnalités BASIS depuis un programme.

- La saisie et validation des documents sera développée à l'aide des modules BASIS (SCREEN et THESAURUS).
- Le système de traitement de textes Q1 pour la saisie de textes.
- La langage C et les modules C-ISAM (gestion des fichiers séquentiel indexés) pour le développement de programmes.
- MFTS ou KERMIT pour le transfert d'information vers la machine centrale.

2. Système central

Hardware

Les ressources à employer peuvent être calculées sur la base de celles qui existent actuellement pour les bases CELEX. Il faudrait y ajouter l'espace pour gérer la BD textes (de l'ordre de 300 MB).

Software

- Le logiciel MISTRAL pour la mise en oeuvre des bases de diffusion (BDD);
- Le logiciel BASIS avec les modules HLI sera employé pour construire la base de données de textes (BDPT);
- Le logiciel de "document processing" MARK-IT sera employé dans le développement du "texterface" pour le traitement et l'analyse des textes provenant des bandes de photocomposition.
- Pour le développement des divers programmes on emploiera C, et les modules C-ISAM et l'interface HLI pour accéder aux bases BASIS.
- Le logiciel de transfert MFTS ou KERMIT

3. Observations

- * Sur le choix du logiciel documentaire pour les bases de diffusion

Dans l'hypothèse d'un choix obligatoire du logiciel, l'option choisie, MISTRAL semble être la plus prudente (voir risques BASIS ci-dessous) parce que c'est le logiciel sur lequel fonctionnent actuellement les bases CELEX, avec des fonctionnalités éprouvées et de bonnes performances au niveau de l'interrogation. Cependant, l'alternative BASIS dans sa version L doit être considérée pour diverses raisons :

- BASIS release L, d'après les spécifications du producteur, peut répondre aux caractéristiques particulières de l'application CELEX, à un niveau de fonctionnalité analogue à celui de MISTRAL (ceci n'est pas vrai pour le release K de BASIS).
- D'un point de vue strictement administratif, BASIS est le seul logiciel documentaire "general use" et, partant, compatible avec la stratégie de la Direction Informatique.
- Par rapport à MISTRAL, BASIS présente certains avantages fonctionnels, citons :

- Gestion plus simple de la base (p.e. plus de problèmes de réorganisation).
- Construction des applications plus simple et en conséquence moins d'efforts de maintenance et de développement.
- Environnement de programmation plus complet (Toolkit, Profile, Report) qui permet d'élargir les fonctionnalités BASIS (p.e. UFI). Avec les modules Profile et Report, même l'utilisateur final pourrait développer ses propres "programmes" d'une façon relativement simple.
- Disponibilité et portabilité sur plusieurs technologies et notamment sur les machines locales.
- Possibilité de stocker dans la base des caractères de contrôle qui ne sont pas visualisés à l'écran.
- Une meilleure perspective de futur, d'après les orientations du producteur (structure de documents, alignement sur la norme ODA, ...).

Cependant le choix de BASIS version L comporte quand même certains risques :

- La version L de BASIS est prévue pour le 1er trimestre 1989 sur les machines VAX et IBM et elle seront disponibles en Europe le deuxième trimestre 1989. Le portage vers SIEMENS et BULL se fera dans le deuxième trimestre de 1990, mais le fournisseur n'a pas d'engagement pour cette date.
- CELEX serait une des premières grosses bases sur BASIS release L. Bien que les releases BASIS soient en général fiables, il y a un certain risque de rencontrer des problèmes inattendus.
- Il reste à vérifier si BASIS est suffisamment performant dans le traitement de textes longs et pour les recherches multichamps. Ceci ne pourra s'évaluer valablement qu'en réalisant et testant un prototype en vraie grandeur.
- Il faut prévoir la formation du "parc" actuel des utilisateurs MISTRAL (+ 1500 utilisateurs) et disposer des outils pour faciliter l'accès aux bases (par exemple, le développement d'une interface conviviale dans le contexte d'IDS).

Tous ces points seront à examiner plus en détail pendant l'étude de faisabilité.

- * Sur le choix de BASIS comme logiciel pour l'archive.

Egalement dans l'hypothèse d'un choix obligatoire du logiciel à ce stade, BASIS a été retenu pour la mise en oeuvre de la base de production (archive) et pour faire la saisie et validation en ligne des documents de l'archive car à l'heure actuelle il constitue l'alternative la plus sûre et éprouvée. Une autre alternative également valable est le SGBDR Oracle et, à ce propos, on peut faire les remarques suivantes :

- Le choix du même logiciel pour le niveau central et local de l'application simplifie globalement le développement et diminue le coût de formation des utilisateurs.
- BASIS et ORACLE sont à l'heure actuelle deux logiciels avec des orientations différentes, et le choix d'un ou l'autre va déterminer la structure finale de l'archive. Pour BASIS, la structure sera orienté "document" et plus proche de celle qui existe dans la base centrale, pour ORACLE la structure sera orientée "record/relation" et on devra faire une décomposition plus fine des champs de l'archive.
- La MAJ dans ORACLE est tout à fait en ligne. Pour BASIS, il faut faire basculer le contenu du fichier Q-File vers la base, bien que pour la saisie et validation, l'utilisateur voie la base comme une seule entité.
- Avec BASIS, tant la structure de la base à définir que le langage d'accès auront une orientation documentaire plus simple et adaptée au contexte actuel des utilisateurs que la structure relationnelle.
- Du point de vue des développements, l'environnement ORACLE semble plus souple, ouvert et complet que celui de BASIS, mais pour le traitement des champs répétitifs et à longueur indéfinie, la solution BASIS est plus simple à mettre en oeuvre. Par contre, pour les tables de validation et de contrôle et la gestion des relations entre les documents, ORACLE est mieux adapté.
- L'évolution future pour les BD relationnelles et les bases documentaires de type textuel semble aller dans le sens d'élargir les fonctionnalités relationnelles avec des fonctions spécifiques de recherche documentaire (p.ex. SQL TEXT RETRIEVAL).

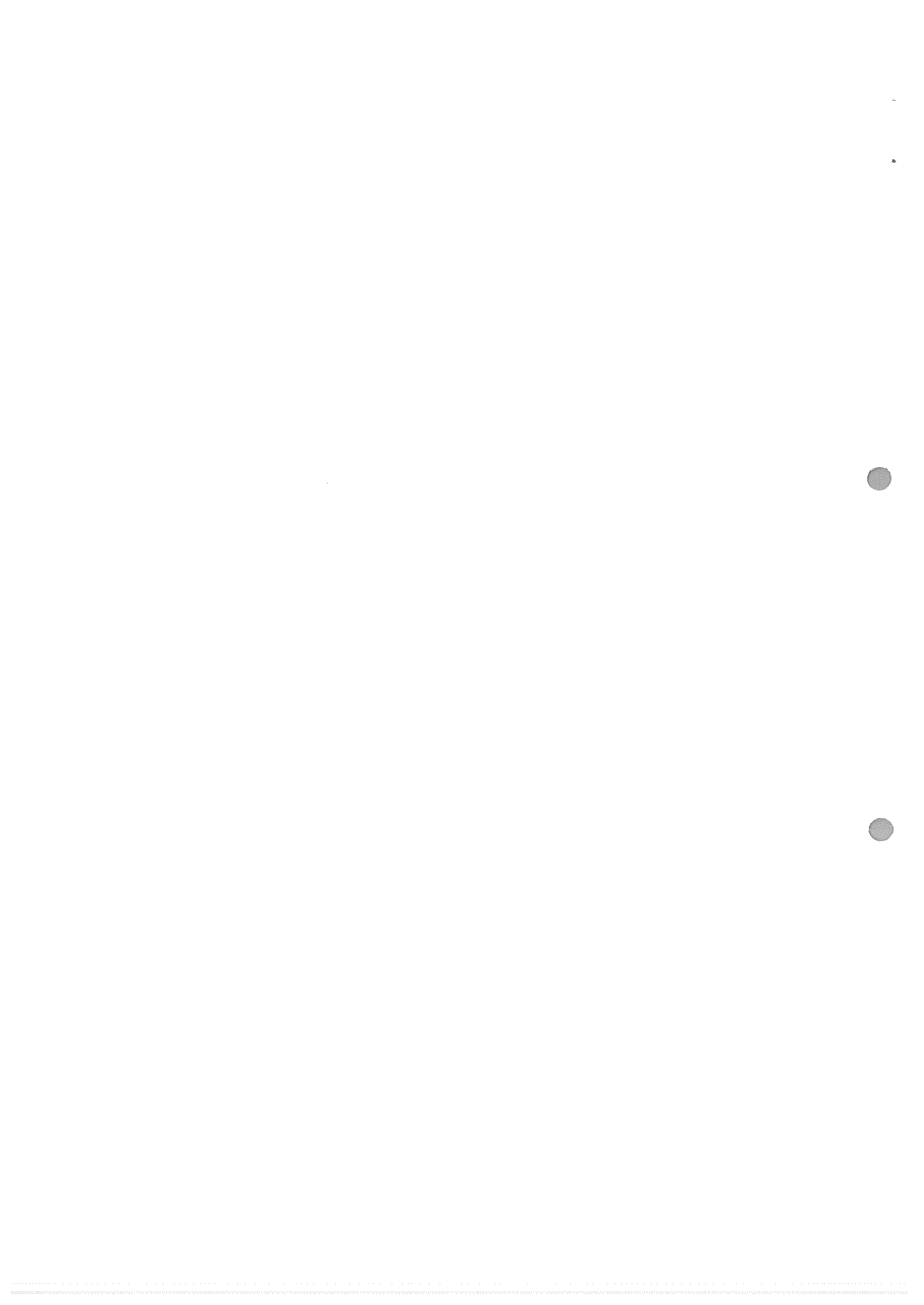
* Sur le choix de BASIS comme logiciel pour la BDPT

Pour l'implémentation de la base de production de textes (BDPT), on peut dire que tant BASIS qu'ORACLE ont le même niveau de fonctionnalité. La raison d'une décision dans le sens BASIS est due à l'intérêt de profiter davantage de l'emploi d'un même logiciel applicatif, l'absence d'expérience ORACLE dans le domaine des bases textuelles et le traitement plus simple des documents longs.

Dans un cas comme dans l'autre, il convient de noter que ce serait le premier développement d'une application de cette taille et avec ces logiciels sur l'ordinateur BULL de la Commission.

* Sur le choix de la machine centrale

Selon les orientations verbales de la Direction Informatique, les machines possibles pour la diffusion des informations sont BULL et SIEMENS. Sous les hypothèses que les logiciels indiqués fonctionnent, le choix est techniquement indifférent et d'autres facteurs (économiques ou politiques) peuvent intervenir, cependant il faut indiquer que sur la machine BULL il n'existe pas de logiciel transfert MFTS et qu'il n'y a pas non plus de prévision de développement. Partant, dans le cas du choix d'une machine BULL, il faudra employer comme logiciel de transfert KERMIT ou faire le développement d'un programme spécifique.



C.2 Relation avec IDS

Une certaine synchronisation doit se faire avec le projet IDS tant dans le cadre des concepts/principes que des outils à mettre en place pour ce projet. Ainsi,

1. Concepts/principes

- Standardisation de la définition des bases de diffusion;
- Le mécanisme de consolidation de la m.a.j. dans la BDD et dans la BDP.
- La définition des critères de cohérence entre la BDD et la BDP, et le mécanisme "d'historisation" des modifications dans la base locale, de façon à les répercuter dans la base centrale.

2. Outils

- L'emploi du logiciel "document processing" MARK-IT pour faire l'analyse, la conversion et le formatage des textes. A ce propos, le développement du "texterface" permettrait de tester valablement MARK-IT.
- Les procédures de transfert et de connexion de la machine locale avec la machine centrale.
- Le mécanisme de down-loading pour extraire un document d'une base centrale, l'envoyer sur la machine locale et construire un fichier textes en format Q-one pour modifier son contenu. Egalement, il doit être possible de reconvertir le fichier Q-one dans un fichier séquentiel qui, par après, sera renvoyé vers la machine centrale.
- Programmes d'interface ISO 2709/SGML pour les logiciels documentaires choisis (BASIS et MISTRAL).
- Interface conviviale (UFI) destinée aux utilisateurs finals. Dans le cas de changement du logiciel de diffusion, l'existence de cette interface permettrait de réduire le coût de formation des utilisateurs.

C.3 Solutions alternatives

Les solutions alternatives suivantes pourraient être envisagées :

1. En ce qui concerne l'architecture de l'application :
installation de la base de production analytique (archive) sur la machine centrale :
Cette solution impliquerait que la production des données (saisie) soit effectuée sur machine centrale, ce qui serait contraire aux principes de développement de l'architecture informatique de la Commission. On pourrait envisager un module de saisie en différé (batch), mais cela conduirait à renoncer aux contrôles en ligne, et donc à une perte de qualité.
2. Installation de la base de production des textes (BDPT) sur machine locale :
Bien que cette solution présente des aspects attirants (production proche de l'équipe DBM, facilité de correction/validation, etc ...), le volume des textes à recevoir (de l'ordre de 300 MB par mois)

nécessiterait une forte augmentation de l'espace disque nécessaire et accroîtrait les coûts de transfert, les risques de problèmes durant les transferts et la dégradation des performances.
L'évaluation de cette alternative sera approfondie durant l'étude de faisabilité du sous-projet "interface textes".

3. En ce qui concerne le choix des logiciels, les différentes possibilités
 - BASIS/L pour les bases diffusion;
 - ORACLE pour les bases de production;ont été discutées dans le chapitre présent.
L'étude de faisabilité se fondera sur l'option MISTRAL pour la diffusion et évaluera de façon exhaustive la solutions ORACLE pour la production.

IV. LE NOUVEAU SYSTEME EN FONCTIONNEMENT COURANT

A. L'ACCES EN LIGNE

Mis à part l'effort qui devra être consenti pour former les utilisateurs aux caractéristiques du nouveau système (voir aussi, ci-dessous, le paragraphe "documentation") deux points méritent d'être signalés dans ce contexte :

- la possibilité d'utiliser le CCL pour interroger CELEX, après que des tests approfondis aient prouvé la faisabilité et le bien-fondé de cette option;
- la mise en place d'une interface conviviale (UFI - voir ci-dessus le point "Relations avec IDS"). Qu'il s'agisse d'un outil général IDS ou d'un développement spécifique à l'application, l'UFI devra être évolutif, et donc aisément paramétrable, dans l'élaboration des stratégies de recherche comme dans le dessin d'écrans.

B. L'UTILISATION EN DIFFERE

Pour CELEX, ce type d'utilisation concerne essentiellement :

- la production du Répertoire de la législation communautaire en vigueur. Il serait utile de revoir la chaîne de production des bandes (éventuellement en utilisant un système de menus), mais aussi les procédures d'édition du répertoire lui-même une fois que CELEX disposera de la richesse typographique nécessaire.
- la production de bandes (livraison initiale et mise à jour périodiques) pour les serveurs externes diffusant CELEX. La chaîne existante pourrait être conservée mais les questions batch correspondant à chaque serveur, ainsi que le lancement de la procédure devraient être sous contrôle du DBM, ici aussi via un système de menus. Cette possibilité économiserait des ressources de gestion (DBA). Par ailleurs, cette économie pourrait encore être accrue par la mise au point - avec toutes les précautions d'usage - d'un système de téléchargement permettant aux serveurs eux-mêmes d'effectuer l'opération de transfert.
- l'exécution de diverses questions batch pour des utilisateurs finals, à titre occasionnel ou périodique. Certains traitements particuliers pourraient être envisagés dans la mesure des possibilités standards offertes par les logiciels utilisés, et sous la responsabilité du DBM.

C. LA MISE A JOUR

La plus grande partie des procédures de mise à jour devra être automatisée et confiée au DBM. Les documents nouveaux et les documents modifiés doivent être aisément repérables dans l'archive, le chargement des textes intervenant après la création de la partie analytique.

Des procédures spéciales sont à prévoir pour :

- la réintroduction de certains champs ou documents via l'interface dans une ou toutes les bases (corrections systématiques);
- la réintroduction de textes après corrections.

Le maintien du strict parallélisme des versions linguistiques entre elles requiert une attention constante de la part de l'équipe DBM. Un traitement des fichiers de production des listings de mise à jour permettrait d'abord de débarrasser ceux-ci des interminables messages-système, incompréhensibles et inutiles, et de disposer d'états synthétiques des mises-à-jour de toutes les versions linguistiques. L'accès en-ligne aux descripteurs nouvellement introduits serait utile.

D. STATISTIQUES

Les contraintes d'administration d'un système tel que CELEX se révèlent particulièrement exigeantes. Les difficultés dues au multilinguisme sont accrues par le grand nombre et la dispersion des utilisateurs.

La restructuration des programmes et procédures de gestion de CELEX doit aller de pair avec une révision des procédures administratives aussi bien que des travaux individuels.

L'administration des bases CELEX, de leurs utilisateurs internes et externes et des serveurs de diffusion nécessite la mise à disposition d'outils convenables. La plupart des opérations traitées aujourd'hui manuellement pourraient être accomplies automatiquement pourvu que la ou les machines disposent des programmes adéquats. L'amélioration de la gestion générale du système suppose une meilleure collecte de données statistiques et l'automatisation maximale des contrôles.

Il va de soi qu'une large part des besoins de CELEX en matière de statistiques sont applicables aux autres bases documentaires gérées par la Commission, y compris dans le contexte UAS. Cependant, on peut admettre que CELEX nécessite des statistiques plus exhaustives, même si celles-ci dépassent les possibilités standard offertes par le logiciel documentaire retenu.

D.1 Statistiques de production

Lors des mises à jour, un certain nombre d'éléments sont nécessaires au suivi de l'archive et des bases comme à la planification des ressources. Les principales données à recueillir sont :

- nombre de nouveaux documents avec taille en octets, par document et par champ; nombre de termes introduits dans les fichiers inverses, avec taille en octets;
- même information pour les documents modifiés;
- répartition par secteurs.

Le nombre de documents nouveaux ou modifiés dans l'archive et les bases doit être identique et automatiquement contrôlé. Sur base de ces données, des statistiques mensuelles et annuelles sont à prévoir.

D.2 Statistiques d'utilisation

Une meilleure perception des utilisateurs et de l'évolution de leurs besoins implique l'existence d'un véritable outil de suivi de leur comportement et de leurs profils d'intérêt; il faut pouvoir déterminer, par utilisateur :

- le nombre d'accès à une base, un lexique, un secteur, ...;
- le nombre d'appels à une procédure particulière;
- le nombre de visualisations d'un champ ou d'un catalogue de champs donné;
- le temps de connexion et le nombre de transactions.

Tout en opérant la distinction entre les utilisateurs des Institutions d'une part et de l'extérieur d'autre part, un système général de suivi de l'accès aux bases permettrait une meilleure gestion des ressources, y compris au niveau de la facturation aux utilisateurs externes.

La diffusion des bases CELEX par des serveurs externes rend nécessaire la mise en oeuvre d'un module spécifique. Les serveurs de la catégorie "gateway" peuvent être assimilés à des utilisateurs en-ligne; par contre, la fourniture hebdomadaire de données sur bandes aux serveurs "normaux" implique qu'un minimum d'informations sur cette procédure soient fournies. Un tel système avait été prévu dans l'analyse du système de fourniture des bandes aux serveurs (module 4) mais n'a pas été développé, étant donné le faible nombre de serveurs, à l'époque, et l'annonce du plan de modernisation. Les informations nécessaires devraient permettre de connaître avec précision quel serveur a reçu quelles données (y compris en volume) et quand.

Plusieurs solutions peuvent être envisagées en fonction de la configuration qui sera définitivement choisie après l'étude de faisabilité. On pourrait par exemple concevoir :

- la création d'une base de données spéciale, sur un SGBD/UNIX, si possible reliée à l'archive et accédant par transfert les données produites sur la machine centrale;
- le téléchargement des données pertinentes sur PC/DOS;
- la création d'une base spécifique sur la machine centrale, pour autant que les outils appropriés soient disponibles.

E. DOCUMENTATION

Au moment de l'ouverture des nouvelles bases CELEX aux utilisateurs internes et externes, les documents suivants devraient être disponibles :

- "Guide d'analyse" (en français)
- "Manuel de gestion", comportant notamment un chapitre "Répertoire des chaînes clientes" (en français)
- "Technical manual for hosts" (en anglais)
- "Manuel utilisateur" (dans toutes les langues pour lesquelles existe une version de CELEX)

A chacun de ces documents doit être associé une procédure rapide de mise à jour et une liste de destinataires, ce qui renforce l'intérêt de pouvoir disposer d'un support bureautique intégré pour la gestion de l'environnement "documentation" de CELEX.

V. DEVELOPPEMENT ET MISE EN PLACE DU NOUVEAU SYSTEME

A. STRATEGIE DE DEVELOPPEMENT

D'après l'architecture de la solution proposée (cf. C.1.A.), le système est composé d'un ensemble de modules qui vont fonctionner autour de trois types de bases de données.

Le projet global de développement sera décomposé en deux sous-projets :

Sous-projet 1 : Production des données textuelles

- construction de la base de production des textes et du "texteface";

Sous-projet 2 : Production des données analyse, diffusion et "pilote"

- construction de la base de production des données analytiques, des bases de diffusion et de l'interface entre production et diffusion;

- construction du module "pilote".

L'intérêt de proposer un projet séparé pour le système d'alimentation des textes s'explique par les éléments suivants :

- on pourrait entrer plus rapidement dans la phase de réalisation concrète du projet de modernisation, tout en se laissant la possibilité d'approfondir d'autres points plus critiques du projet.

- ce développement peut s'intégrer sans difficultés techniques dans le mode de fonctionnement actuel de CELEX, qu'il contribuerait à rationaliser en remplaçant les chaînes "textes" actuelles, complexes, coûteuses en ressources et génératrices de problèmes. Il permettrait par ailleurs l'enrichissement des textes des bases de diffusion actuelles.

Cette stratégie nécessiterait le développement d'une interface entre la base production des textes et MISTRAL V5, actuellement utilisé pour les bases de diffusion.

La séquence de développement des autres modules peut s'envisager de la manière suivante :

- saisie/alimentation et construction de la base de production analytique (archive);

- interface locale;

- construction des bases de diffusion;

- interface centrale;

- module de diffusion vers les sites serveurs externes;

- module pilote.

Il convient de noter que, d'un point de vue strictement technique, il serait possible de développer ces modules séparément puis de procéder à leur intégration et éventuelle mise en production. La construction des interfaces de transfert implique cependant la définition préalable des paramètres des bases de diffusion.

B. STRATEGIE DE MISE EN PLACE (Scénario de mise en place)

CELEX est une application dont le fonctionnement implique de nombreux intervenants et qui est diffusée à un grand nombre d'utilisateurs pour lesquels elle constitue souvent un outil de travail. Il faut pouvoir mettre en place le nouveau système sans que le service fourni à ces utilisateurs ne soit pénalisé.

Dans le but d'exprimer d'une façon claire et suffisante précise le contexte pour la mise en production du projet de modernisation CLX, on va employer un modèle simple (fig. B.1) du fonctionnement actuel de l'application ainsi qu'un modèle de la situation finale selon la solution proposée (fig. B.2).

Sur la base de ces deux schémas on peut dire que le problème à résoudre est : trouver la meilleure méthode (ou stratégie) de passer de la situation de la fig. B.1 à la situation de la fig. B.2.

D'accord avec la stratégie de développement (S V.A) discutée, le projet global peut démarrer avec la réalisation du sous-projet "texterface" et en conséquence avec les actions suivantes :

- construction de BDPT (textes) et du "texterface";
- mise en production de ces deux éléments et intégration avec le système actuelle;
- poursuite de la production courante, en utilisant ce développement au lieu des chaînes textes actuelles.

Avec des action on évolue du schéma fig. B.1 à celui de la fig. B.3.

A ce point-ci on a plusieurs alternatives basées sur la façon dont les modules développés sont mis en production, soit tous les modules à la fin du projet, soit d'une façon progressive selon un ordre de développement défini. Parmi un éventail possible, on va discuter les " solutions les plus "faisables".

Alternative 1 : Mise en production du projet complet.

La séquence des actions à suivre est la suivante :

A. Partie "analyse"

1. Génération de la nouvelle BDPA (archive) et chargement de celle-ci à partir de l'archive actuelle;
2. Chargement initial des nouvelles bases de diffusion à partir de la BDPA (nouvelle archive);
3. Récupération des données des MAJ archive et bases (ancien système) qui ont eu lieu pendant les actions 1. et 2. et MAJ de la BDPA (nouvelle archive) et des nouvelles bases. Cette action sera répétée jusqu'à ce que les deux archives (nouvelle et ancienne) et les bases (seulement leurs parties analyses) soient au même niveau.

4. A chaque MAJ de l'archive et des bases (ancien système), les données "analyses" seront transférées vers le nouveau système pour y faire la MAJ de la BDPA et des bases. Le processus devra fonctionner jusqu'à l'abandon du système ancien.

A remarquer que la saisie des données analyse se fera toujours sur l'ancien système.

B. Partie "textes"

Au moment que les nouvelles bases sont définies, on peut réaliser les actions suivantes :

1. Pour maintenir le parallélisme, les nouveaux textes à charger dans les mises à jour courantes leurs seront à la fois dans les anciennes et les nouvelles bases de diffusion.
2. Les textes ayant déjà fait l'objet d'un chargement dans les anciennes bases, via la base de production des textes, seront rechargés dans les nouvelles par le même module (BDPT + "texterface").
3. Pour les textes existant dans les bases actuelles de CELEX mais non sur la base de production des textes, un processus de transfert analogue à celui utilisé par la diffusion vers les serveurs externes (ISO 2709) sera utilisé.

Le schéma de la fig. B.4.1. représente la situation décrite et la fig. B.4.2. les actions dans le temps.

Alternative 2 : Mise en production du module BDPA (archive) avant la fin du projet.

A. Partie "analyses"

1. Génération de la nouvelle BDPA (archive) et chargement de celle-ci à partir de l'archive actuelle.
2. Récupération des données de la MAJ archive et bases (ancien système) qui ont eu lieu pendant le déroulement de l'action 1. et réalisation de la MAJ dans le nouveau système (archive).
3. Une fois que les deux archives sont à niveau, la saisie de la partie analyse se fait sur le nouveau système avec l'abandon de l'ancien.
4. Pour la MAJ archive et bases (ancien système), les données sont saisies sur la BDPA (nouvel archive) et converties dans le format chaîne d'entrée de l'ancien système. Ce processus sera itéré jusqu'au développement des modules interface et BDD's du nouveau système.
5. Génération des nouvelles bases de diffusion à partir de la nouvelle archive jusqu'à ce qu'elles soient à niveau dans leurs parties analyse.
6. Saisie des données sur la nouvelle archive, MAJ des nouvelles bases, conversion de données en format chaîne d'entrée et MAJ archive et bases (ancien système). Ce processus devra fonctionner jusqu'à l'abandon du système ancien.

B. Partie "textes"

Le processus à suivre est le même que celui de l'alternative 2.

Le schéma de la fig. B.5.1. et B.5.2. représentent la situation décrite et le B.5.3. les actions dans le temps.

Alternative 3 : Mise en production des modules BDPA (archive) et interface avant la fin du projet

La séquence des actions est la suivante :

A. Partie analyse

1. Génération de la nouvelle BDPA (archive) et chargement de celle-ci à partir de l'archive actuelle.
2. Récupération des données de la MAJ archive et bases (ancien système) qui ont eu lieu pendant le déroulement de l'action 1. et réalisation de la MAJ dans le nouveau système (archive).
3. Une fois que les deux archives sont à niveau, on fait la saisie sur le nouveau système et on abandonne la chaîne analyse du système ancien y inclus l'archive. Les données analytiques pour les bases anciennes seront produites par un programme interface ad-hoc (interface "old analyse").
4. Saisie des données sur la nouvelle archive, génération données "analyse" pour les anciennes bases et MAJ de ces bases. Ce processus sera itéré jusqu'à la mise en place des nouvelles bases et le chargement initial de celles-ci.
5. Avec les bases anciennes et nouvelles à niveau dans leurs parties analyse et processus à suivre sera : saisie des données dans la nouvelle archive, MAJ nouvelles bases, conversion données format anciennes bases et MAJ de celles-ci.

B. Partie textes

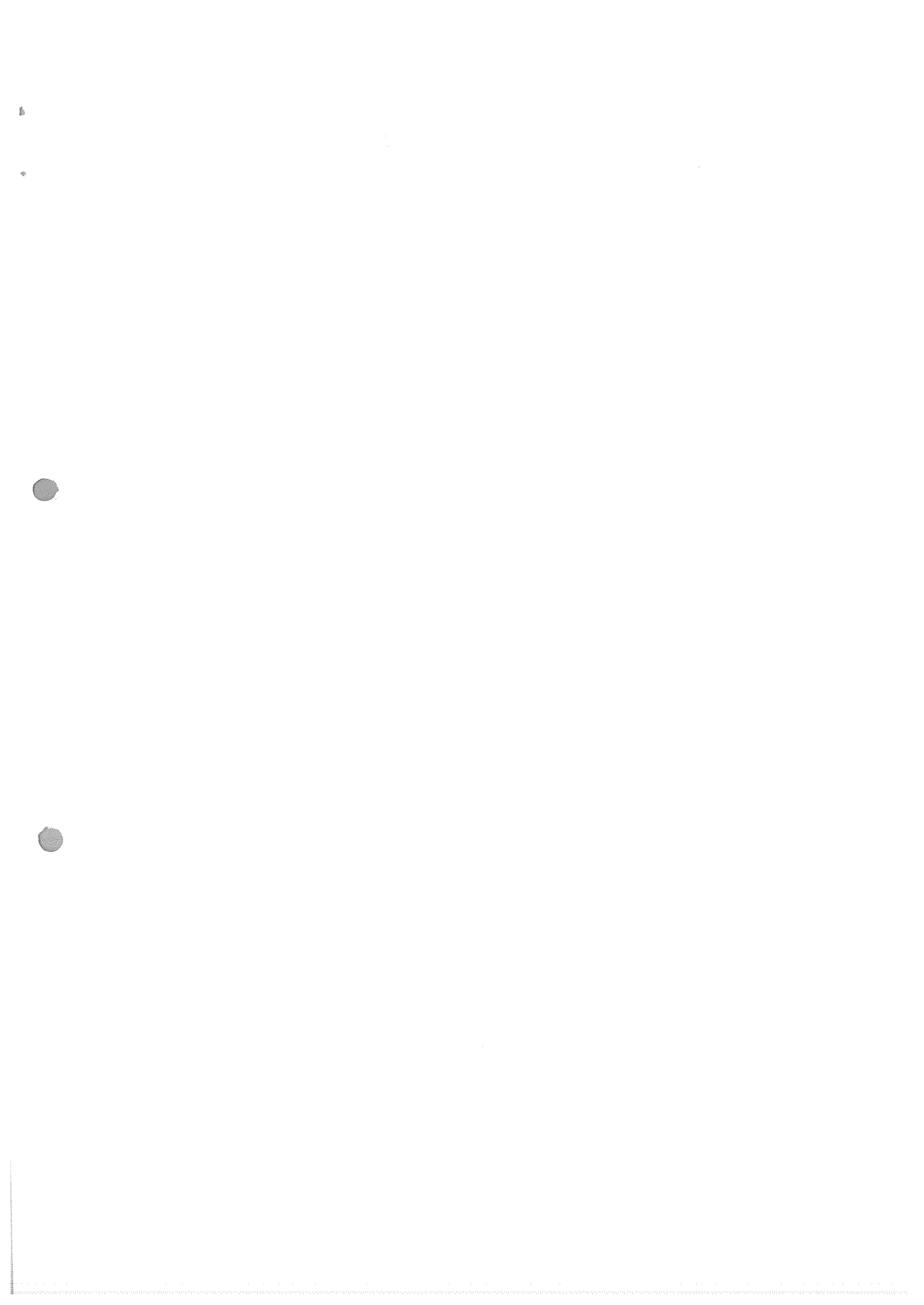
Le processus à suivre est le même que celui de l'alternative 1.2.

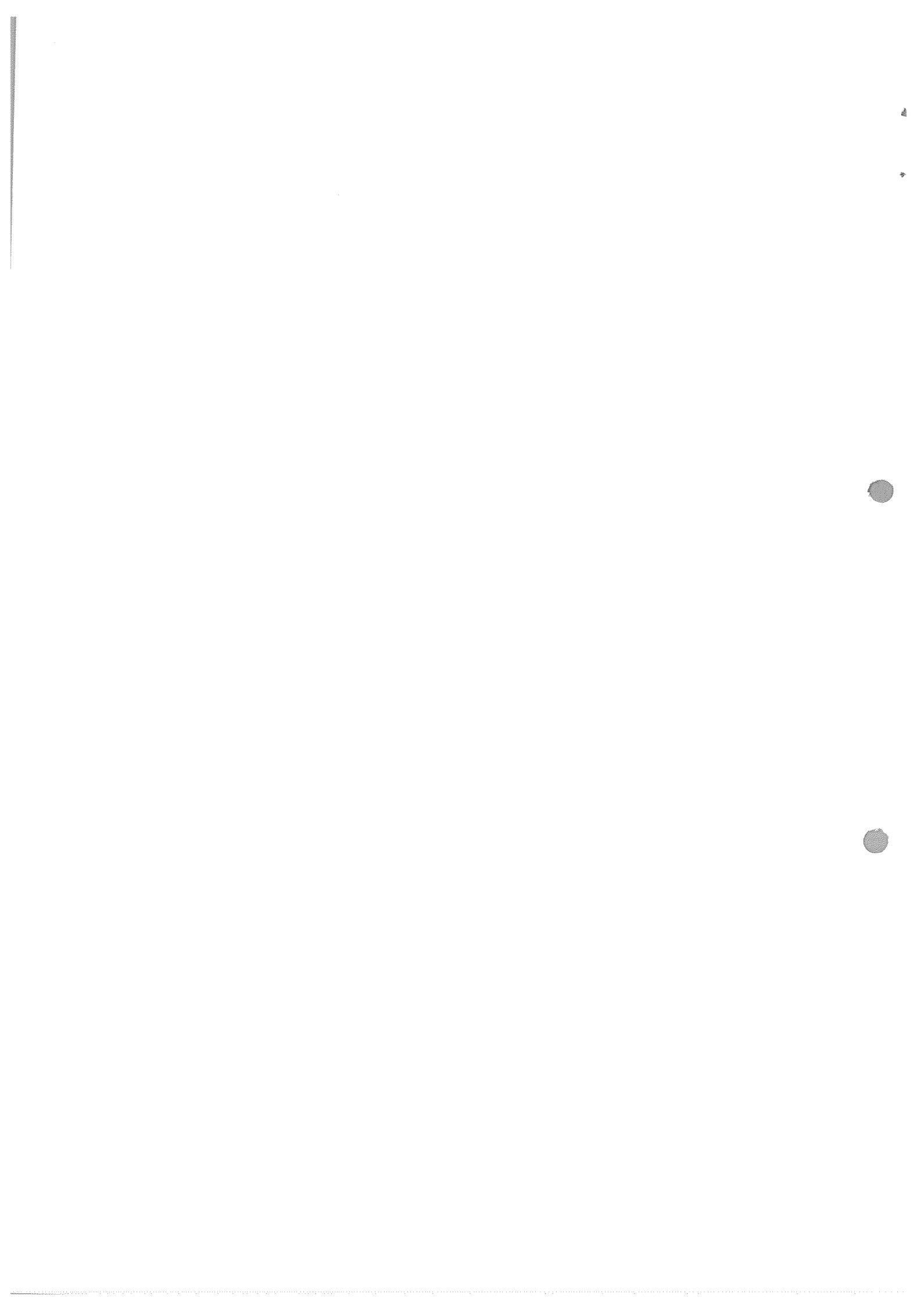
Le schéma des fig. B.6.1., B.6.2. représente la situation décrite et le B.6.3. les actions dans le temps.

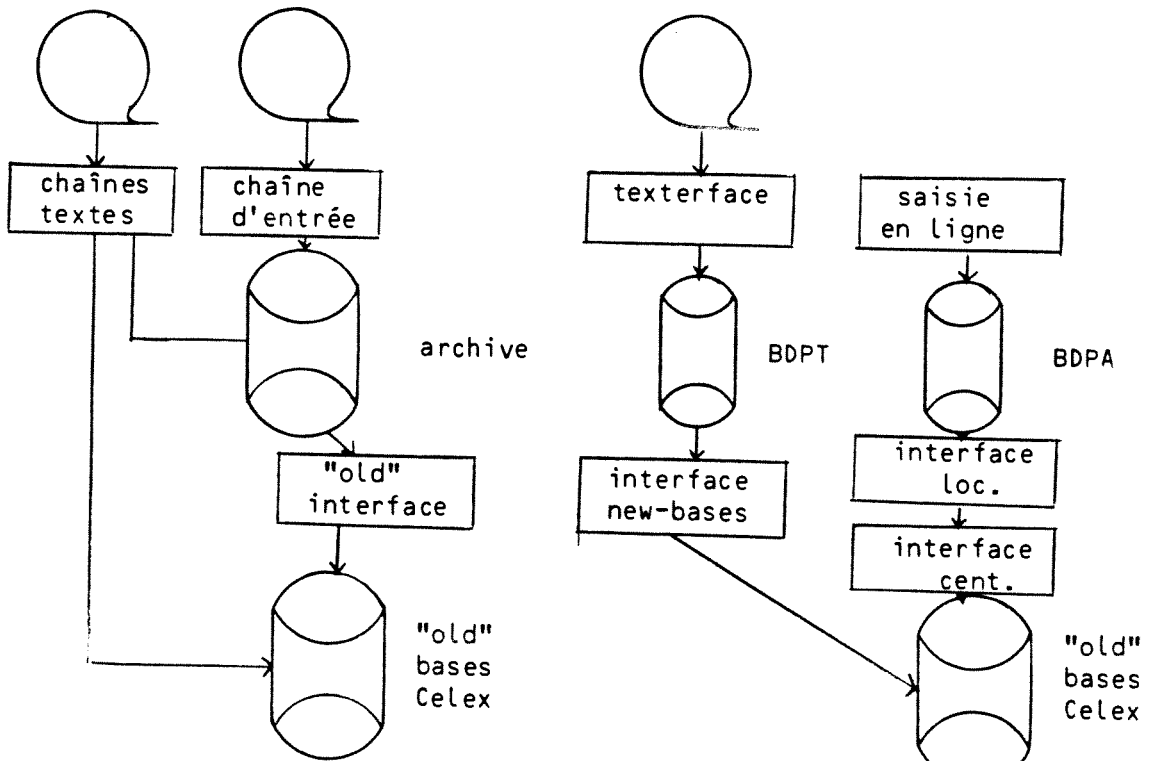
Observations

- L'alternative 1 est celle qui a moins d'intégration avec le système actuel, en fait il y en a aucune, étant donnée que le processus de production actuelle reste inchangée. Par contre, elle présente une lourde charge de gestion, il y aura deux systèmes en parallèle.

- On peut remarquer dans les trois alternatives que le processus de mise en place des modules développés et le rattrapage des textes constituent en eux-mêmes une sorte de "parallèle". Ceci veut dire qu'un processus du "parallel running" classic pour ce projet n'a pas beaucoup de sens sauf pour les utilisateurs finaux.

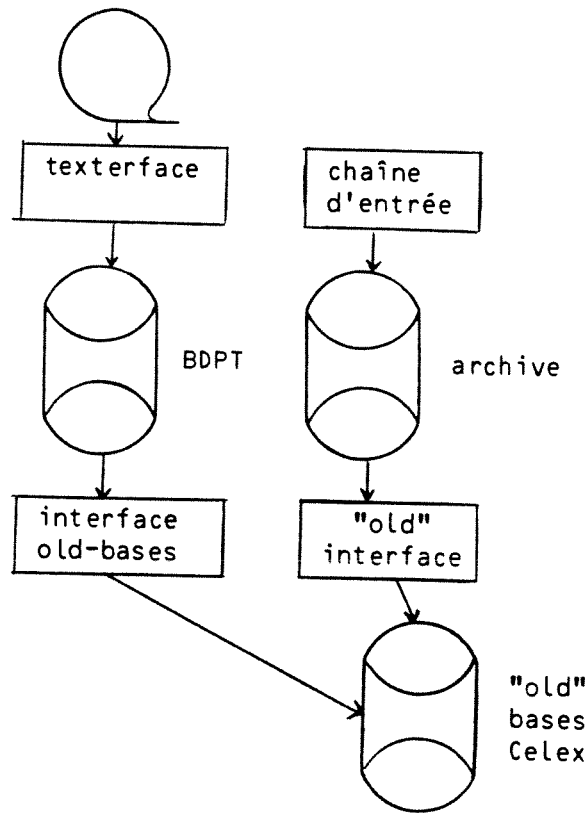




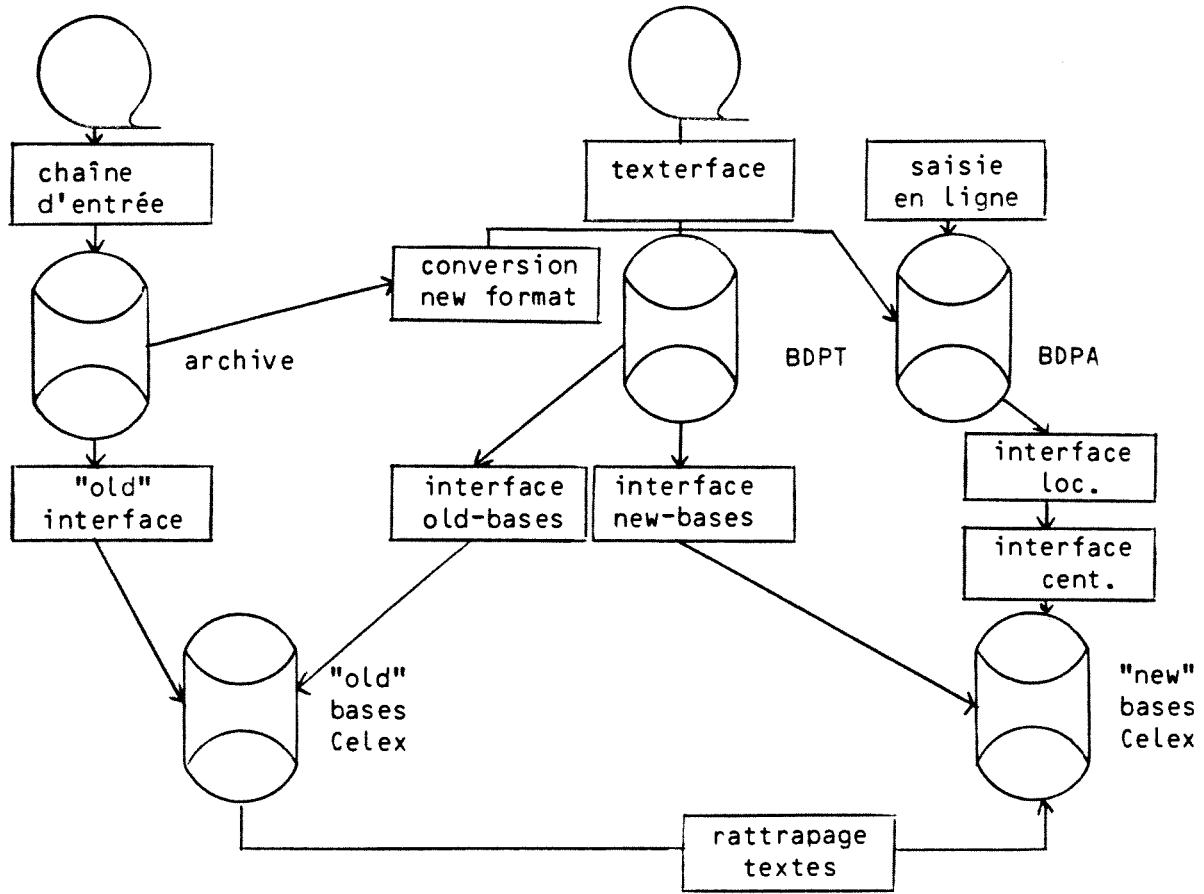


Situation actuelle (fig. B.1)

Situation finale (fig. B.2)



Situation après sup texterface (fig. B.3.)



Situation alternative 1 (fig. B.4.1)

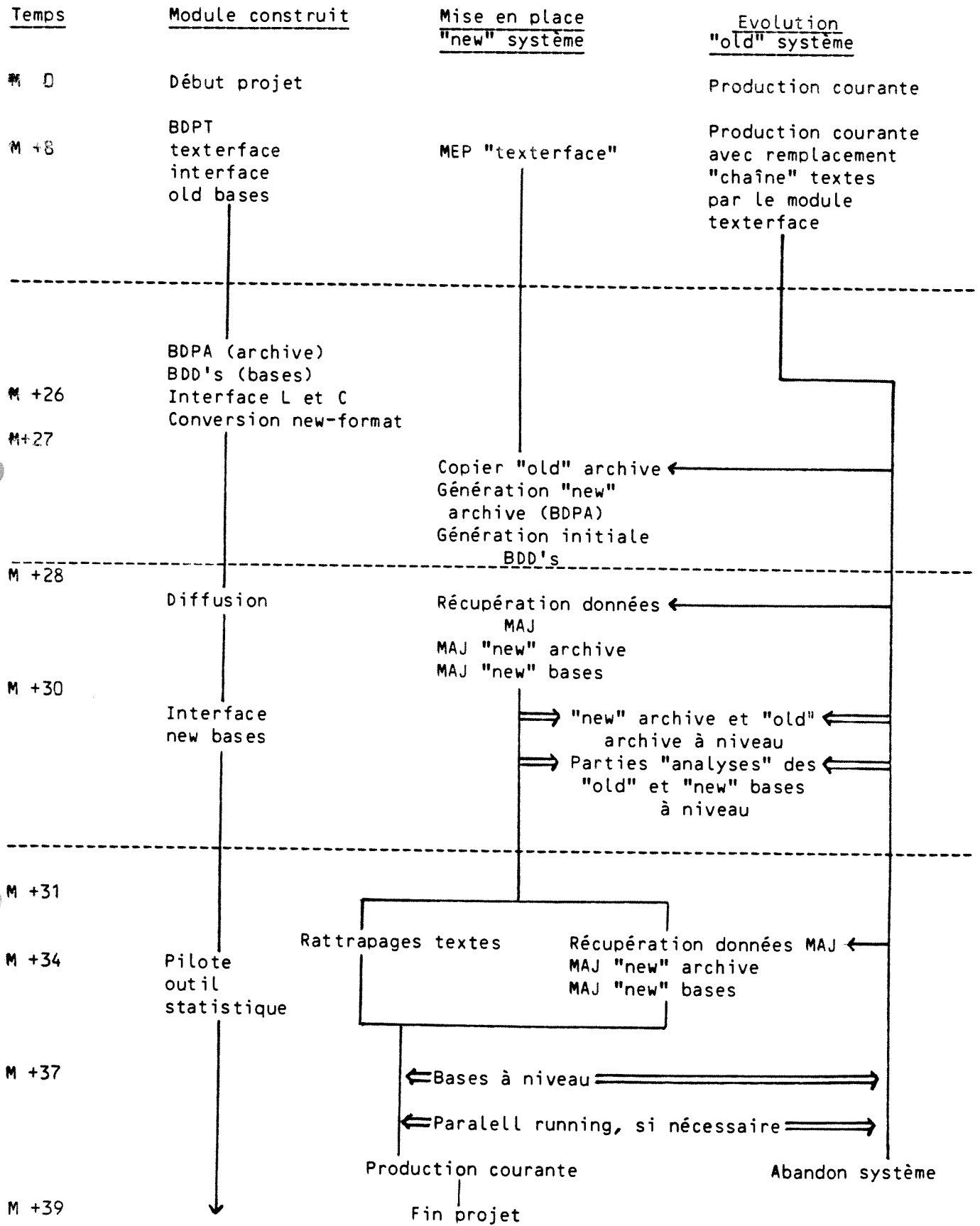
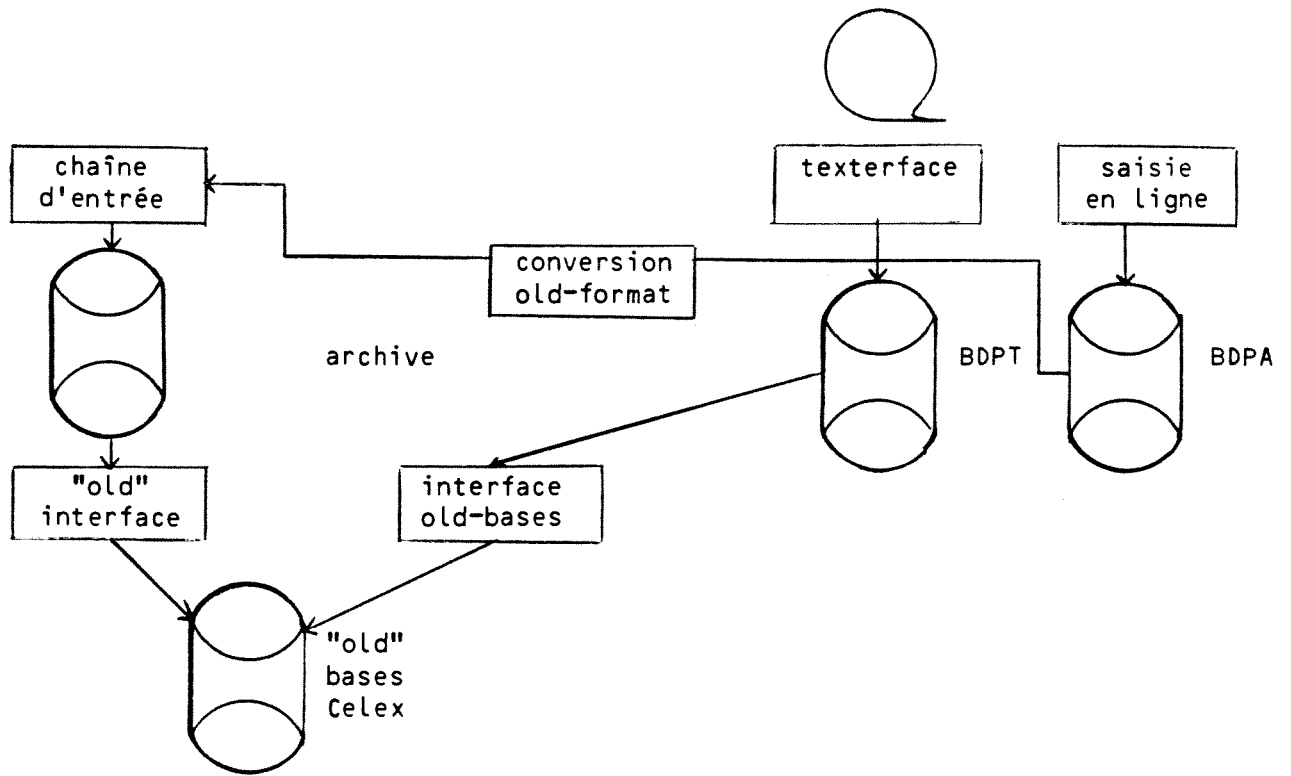
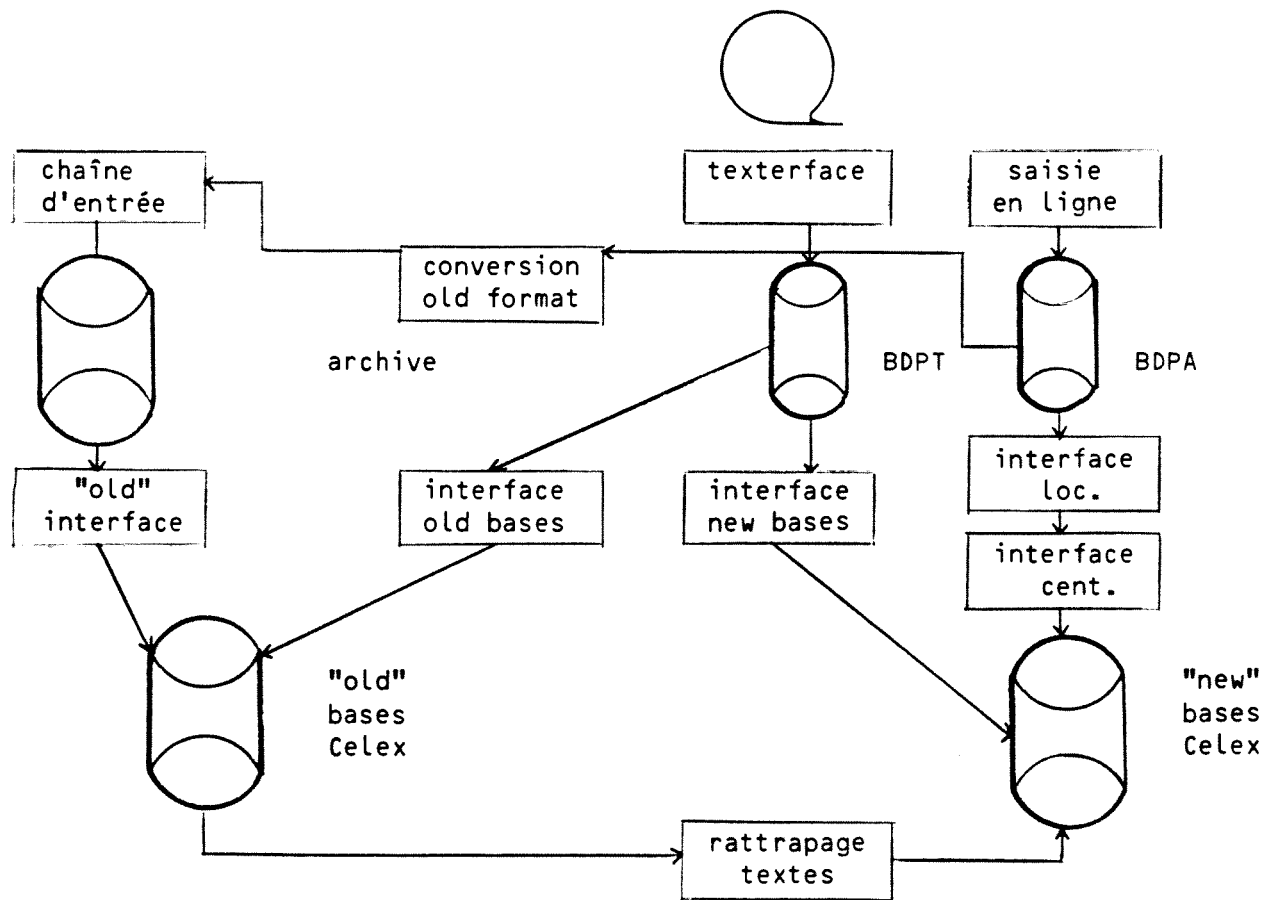


Fig. B.4.2 (Alternative 1)



situation alternative 2 (fig. B.5.1.)



Situation alternative 2 (fig. B.5.2)

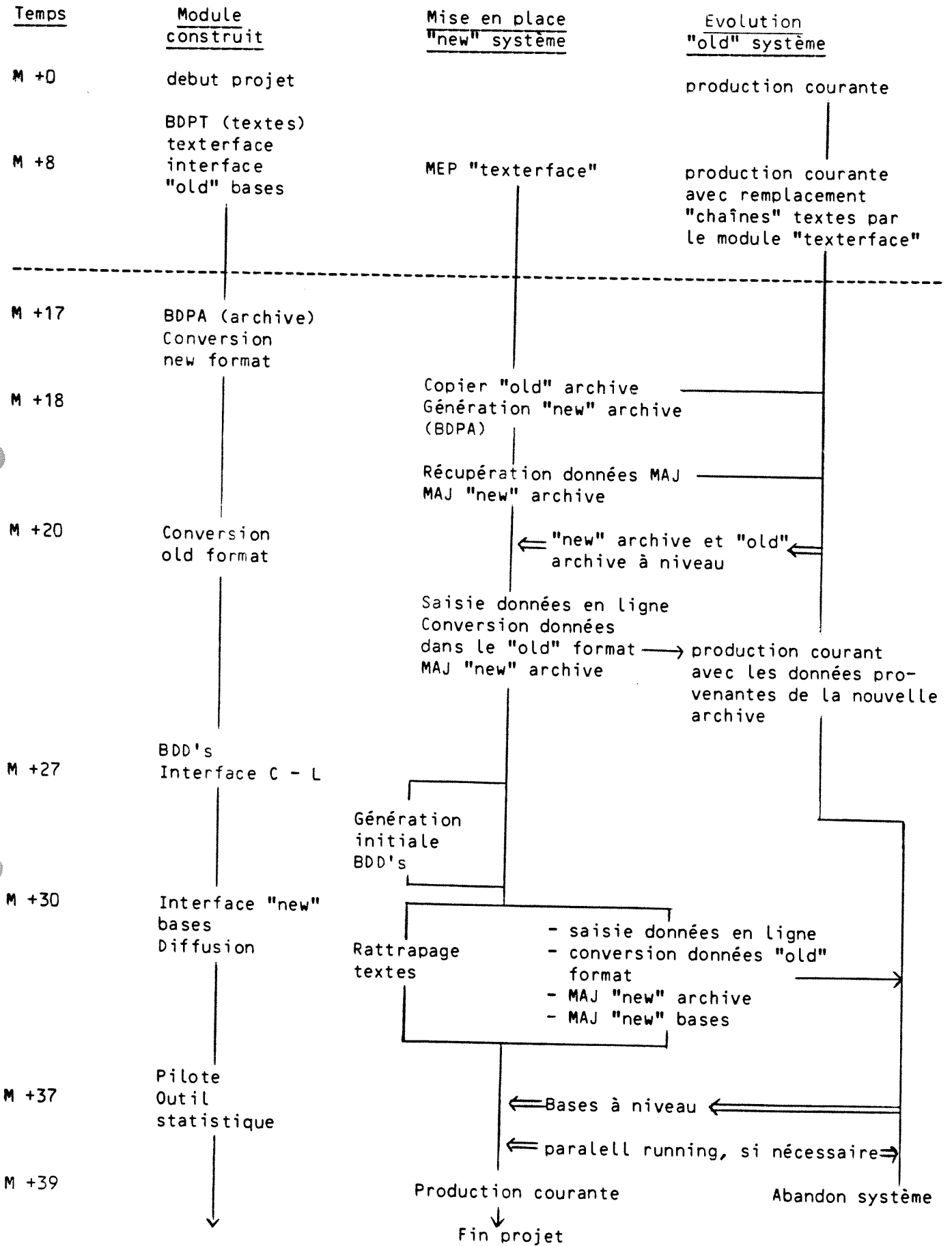
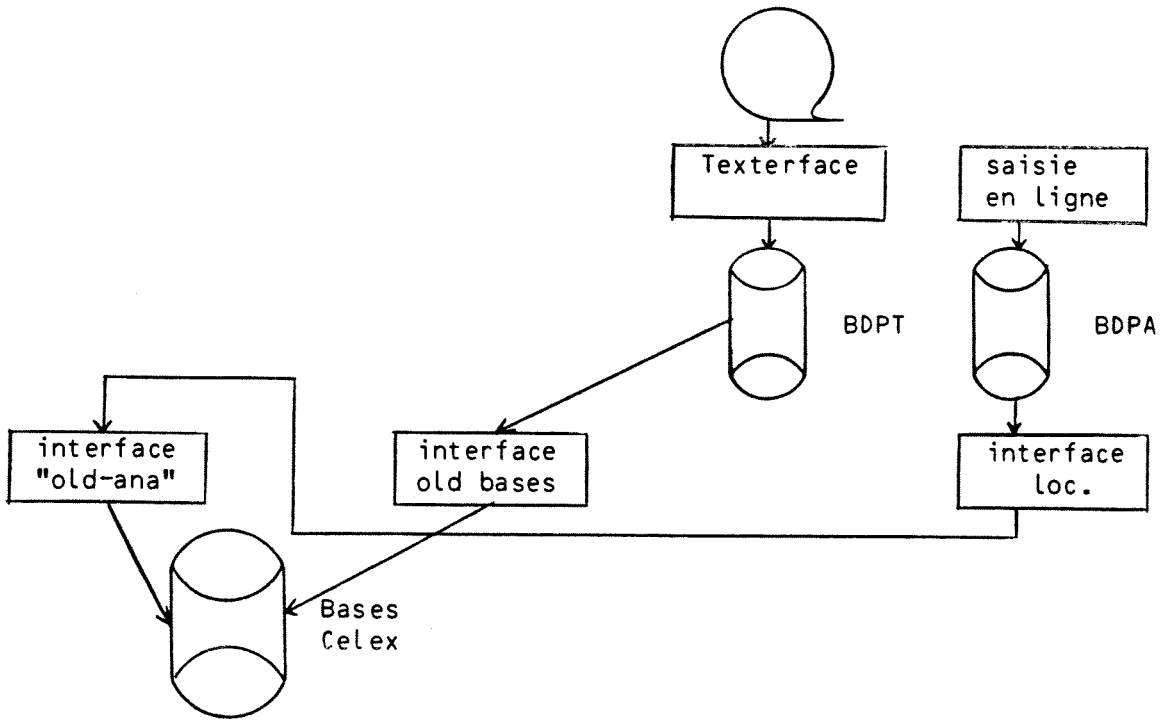
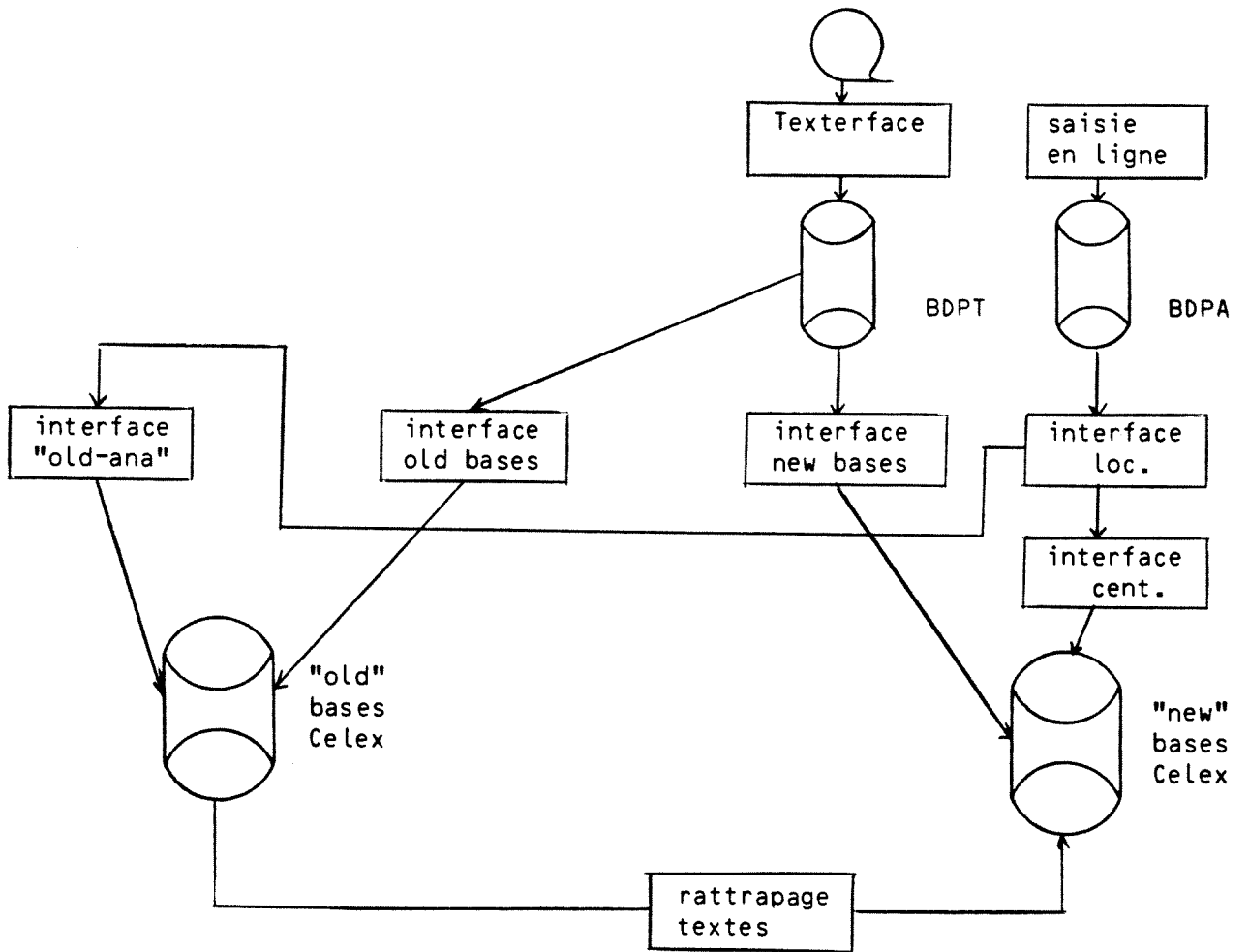


Fig. B.5.3 (Alternative 2)



Situation alternative 3 (fig. B.6.1)



Situation alternative 3 (fig. B.6.2.)

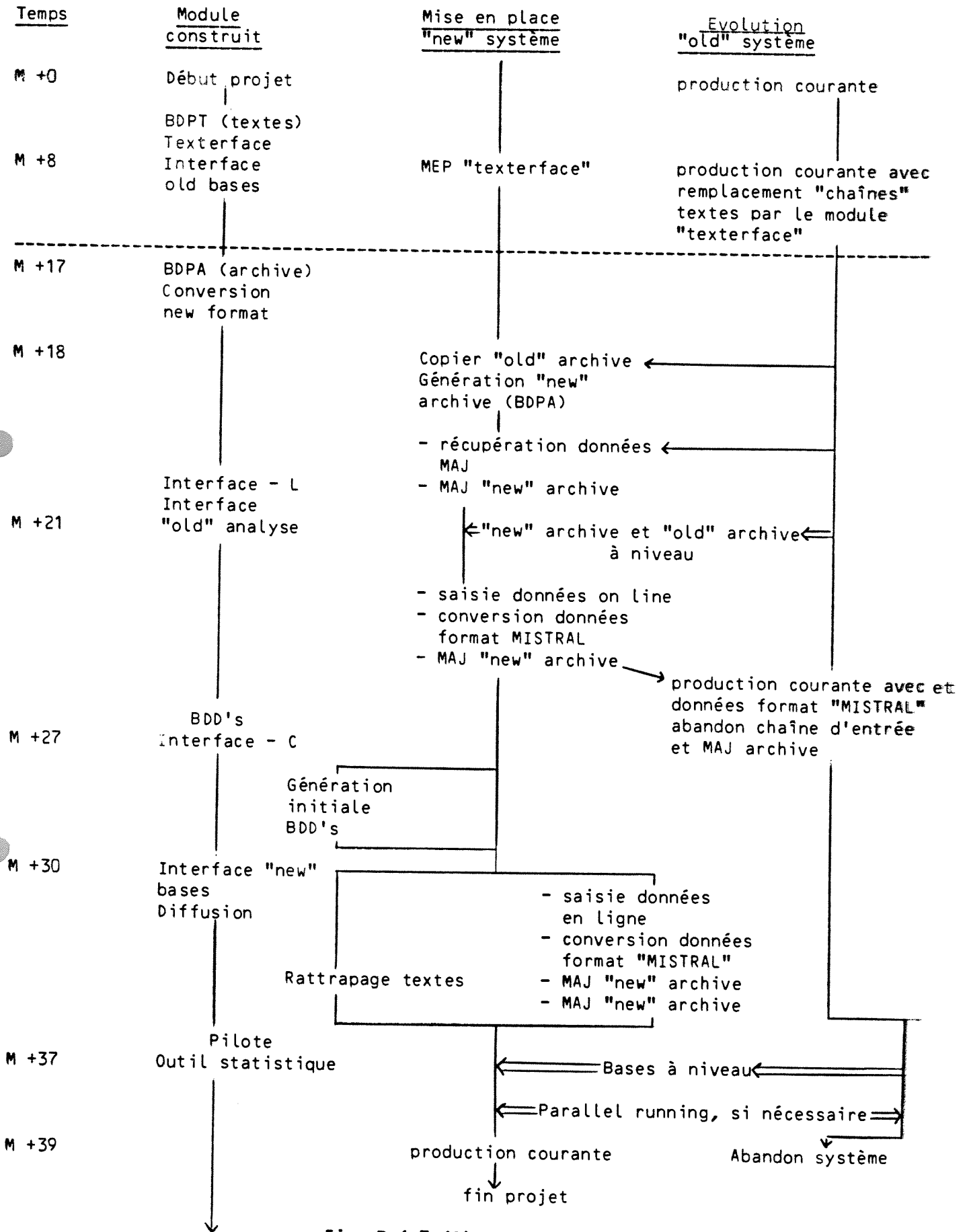


Fig. B.6.3 (Alternative 3)

C. ESTIMATION DES RESSOURCES

Sur la base de la description fonctionnelle et de la structuration en modules proposée, on peut avancer les ressources (exprimées en homme/mois) nécessaires pour le déroulement du projet. Au stade de la préanalyse, il s'agit de chiffres indicatifs devant être confirmés par l'étude de faisabilité.

	SDA	SII dev	CC gestion	CC DBSA
<u>Etude de faisabilité</u>				
- Validation BASIS sur GCOS 8				3
- E.F. texterface et BDPT (sous-projet 1)	3	1	-	
- E.F. construction BDD, interface (sous-projet 2)	6	4	-	
Total :	9	5	-	

Construction

Sous projet 1 :

- Texterface	3	1	-	
- Construction BDPT	1	0,5	-	
- Interface BDPT-BDD	1	0,5	-	
- Test/intégration CELEX actuel	2	1	2	
Sous-total 1 :	7	3	2	

Sous projet 2 :

- Saisie/alimentation	3	1	-	
- Construction archive	2,5	1	-	
- Interface locale	2	0,5	-	
- Interface centrale	3	0,5	-	
- Construction BDD's	3	1	-	
- Test/intégration BDPA et BDP	2	1	2	
- Diffusion	1	0,5	0,5	
- Outil statistique	2,5	1	-	
- Pilote	4	1	0,5	
Sous-total 2 :	23	7,5	3	
Total :	30	10,5	5	

Génération/chargement nouvelles bases

- Outil de conversion de l'archive	2	1	2	
- Outil de comparaison old/new CELEX	1	0,5	-	
- Rattrapage des textes	1	2	6	
- Outil de transfert des données de la nouvelle archive vers l'ancien système (base et/ou archive) d'entrée	3	1	2	
Total :	4	3,5	7	

Parallel running

Total :	1	2	3	
----------------	----------	----------	----------	--

=====
TOTAL GENERAL : 47 22 18 3



VI. ETUDE DE FAISABILITE

Afin de pouvoir déterminer avec le maximum de précision les coûts et les bénéfices à attendre de la nouvelle application (construction et fonctionnement courant), ainsi que d'effectuer convenablement la réalisation, les textes et la mise en production de celle-ci, l'étude de faisabilité devra fournir les spécifications détaillées des différents modules. Les actions relevant de l'étude de faisabilité propre à chacun des sous-projets proposés sont les suivantes :

Sous-projet 1 : système de production des données textuelles

1. évaluation/décision sur l'architecture de l'application (système central ou local);
2. spécifications de la base de production des textes (BDPT);
3. validation du logiciel à employer (ORACLE ou BASIS) à l'aide d'un prototype;
4. analyse des différents formats d'entrée (FORMEX/SGML, IC2, ...);
5. spécification du système d'analyse des documents avec :
 - structuration des textes selon les champs logiques CELEX;
 - production/initialisation des bordereaux d'analyse;
 - traitement du multilinguisme (ISO 8879/1, ISO 8879/7, ISO 2022);
 - outils de transcodage, translittération et appauvrissement;
 - outils de correction/modification des textes par l'équipe DBM;
 - spécification de l'interface avec les bases de diffusion actuelles;
 - plan de rechargement et enrichissement des bases.

Sous-projet 2 : Système de production des données analytiques, interface, diffusion et pilotage

- A. Base de données de production (Archive)
 - définition de la structure de la base;
 - évaluation/validation du logiciel à employer (BASIS ou ORACLE);
 - système d'analyse en ligne;
 - gestion des annulations;
 - gestion des tables de validation et contrôle;
 - outils de gestion de la base
- B. Interface locale et centrale
 - structure de document-pivot CELEX;
 - spécification des règles de transformation des données de la base de production dans les bases de diffusion (tables de correspondance);
 - relations avec la base de production des textes
- C. Bases de diffusion
 - définition;
 - spécification des outils de gestion;
 - spécification des outils de diffusion aux serveurs externes (production des bandes ISO 2709, gestion des serveurs ...);
 - construction d'un prototype destiné aux utilisateurs finals
- D. Module "pilote" et outils statistiques
 - spécification des fonctionnalités du module "pilote" et de l'interface DBM;
 - outils de suivi des statistiques de production (textes et données analytiques ...);
 - production des statistiques d'utilisation des bases de diffusion

E. Mise en place du nouveau système

- consolidation de la stratégie proposée;
- spécification du processus de génération de la nouvelle base de production (archive);
- spécification du processus de génération des nouvelles bases de diffusion

F. Divers

- structure du nouveau secteur 7;
- rédaction du manuel utilisateur final (liée à la mise en place du prototype)

VII. COÛTS ET BÉNÉFICES

Il serait hasardeux, dans la phase préliminaire d'un projet d'une telle envergure, de vouloir procéder à une analyse détaillée et exhaustive des éléments économiques. Des estimations raisonnables peuvent cependant s'envisager.

A. LES COÛTS

Les **coûts de construction** ont fait l'objet d'une première évaluation (cf. point V.C ci-dessus). Cette estimation (+ 7h/an) devra être précisée et confirmée dans l'étude de faisabilité.

Les **coûts en matériel** seront essentiellement dûs à l'installation de la machine locale et de ses périphériques. L'espace disque occupé, les temps CPU et réseau seront semblables à ceux du système actuel (au moins globalement) et évolueront de façon plus organisée que ce n'est le cas maintenant, y compris en ce qui concerne les coûts relevant de l'extension de la couverture (cf. point III.A.6 ci-dessus).

Les **coûts opérationnels** diminueront sensiblement, surtout en ce qui concerne les ressources humaines. On peut déjà penser que la maintenance du système ne nécessitera plus qu'une personne (au lieu de deux), la gestion, une personne (au lieu de 2,5) et l'encodage deux personnes (au lieu de quatre, sans compter les 140 000 ECUS consacrés annuellement à l'encodage externe (poste 2130)).

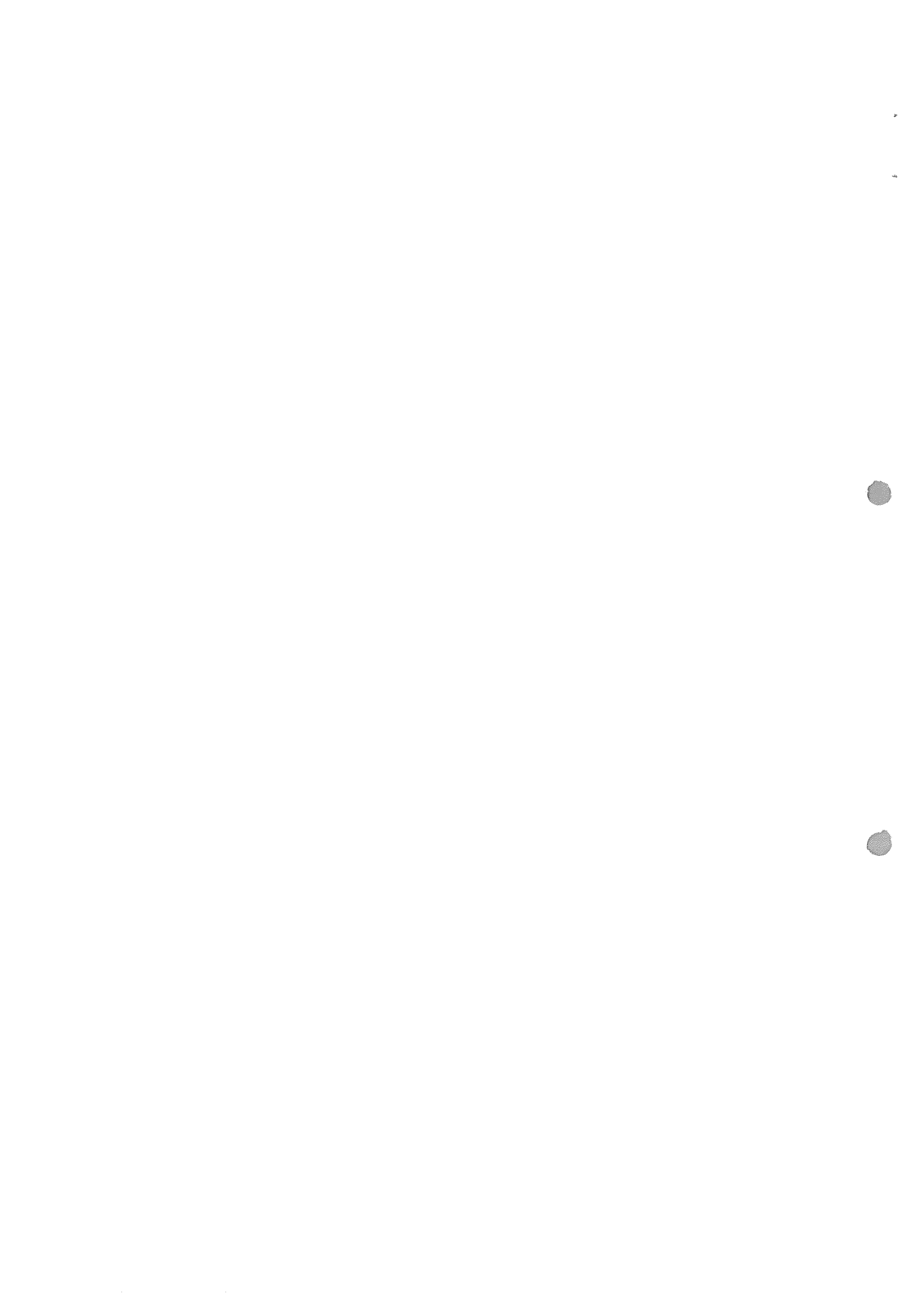
Les **coûts d'administration** devraient rester stables, mais l'accroissement de la productivité générale devrait permettre une nouvelle répartition des fonctions au sein de l'équipe DBM pour mieux faire face à l'arrivée des nouvelles versions linguistiques et d'une population croissante d'utilisateurs, dont la formation et le suivi doit s'effectuer de façon plus soutenue que ce n'est le cas actuellement.

Les **coûts d'utilisation**, enfin, devraient diminuer en fonction des nouvelles possibilités d'accès plus rapide à une information plus complète et **directement utilisable** (meilleure présentation et texte intégral en riche).

B. LES BÉNÉFICES

Les bénéfices attendus du nouveau système sont de tous ordres et concernent tous les intervenants, depuis le SDA jusqu'à l'utilisateur final. Outre les réductions de coûts mentionnées ci-dessus, un grand nombre de ces bénéfices sont très difficilement quantifiables, surtout à ce stade du projet. Vouloir les énumérer aboutirait en fait à une répétition des objectifs décrits dans le point II du présent document.

On pourrait par ailleurs se demander si les Institutions, et notamment la Commission, devraient subordonner à une pure relation arithmétique entre les coûts et les bénéfices la mise à disposition du public européen d'un tel instrument d'information à l'horizon 1992.



Annex 1. PRESENT SITUATION



Annex 1A. Multilingual fields and lexica

T A B L E A U D E S L E X I Q U E S E T D E S C H A M P S

CLX	CDE	CEN	CML	CIT	CDA
BIBLIO num.doc auteur forme traité typ.doc	BIBLIO dok.num autor form vertrag dok.typ	BIBLIO doc.num author form treaty typ.doc	BIBLIO number auteur vorm verdrag reeks	BIBLIO numero autore forma trattato serie	BIBLIO doknum udsteder form traktat doktyp
REF.PUB	FUNDST	PUB.REF	VINDPL	FONTE	PUBREF
JO titre texte texte02-09 rectif rectif02-04	ABL titel text text02-09 berichtg bericht02-04	OJ title text text02-09 corrig corrig02-04	PB titel tekst tekst02-09 rectif rectif02-04	GU titolo testo testo02-09 rettif rettif02-04	EFT titel tekst tekst02-09 berigt berigt02-04
RECUEIL tit.jur index sommaire intit.1 intit.2 cour motifs motifs02-09 visa depens disposi	SAMPLUNG fall.tit schlagw leitsatz gegenst1 gegenst2 richter gruende gruend02-09 bezeuge kosten tenor	REPORTS case.tit index summary issue.1 issue.2 court grounds ground02-09 recitals costs op.part	JURISPR zaak trefw samenv voorw.1 voorw.2 hof r.overw r.over02-09 aanh kosten dictum	RACCOLTA causa epigrafe massima oggetto1 oggetto2 corte motiv motiv02-09 richiami spese disposi	SML sagtit stikord smndrag genst01 genst02 domst praem praem02-09 omtbest sagsomk konkl

INDEX.PE
descript
resumé

DATE
pub
doc
vote
dem
notif
envoi
effet
échéance
débat
fin.val
réponse
transpos
ratif
signat

INDEX.EP
descript
abstract

DATUM
veroeff
dok
abstimmg
eingang
zustellg
absend
inkraft
termin
sitzung
end.guel
antwort
umsetzg
ratif
untzeich

INDEX.EP
descript
abstract

DATE
pub
doc
vote
lodged
notified
despatch
inforce
deadl
debate
end.val
reply
transpos
ratified
signed

INDEX.EP
descript
resumé

DATUM
pub
doc
stemming
inschr
kennissg
verzend
werking
vervald
debat
eind.gel
antwoord
omzett
ratif
ondertek

INDEX.PE
descript
riass

DATA
pub
doc
voto
registr
notif
trasmis
effetto
scadenza
discuss
fineval
risposta
recepti
ratifi
firma

INDEX.EP
descript
resume

DATO
offgoer
dok
afstem
indg
medd
afs
lraft
frist
forh
ophoer
svar
gnmfoer
ratif
underteg

REL
bas-jur
cit.div
act.mod
mod.ult
cit-jur
act-jur
act.vis
vis.ult
arrets
ant-jur
ult-jur

ZIT
grundl
zit.akte
aendert
geaendrt
jur.zit
betrifft
bezugn
spvorarb
urteile
fr.urt
sp.urt

CROSSREF
leg.bas
leg.cit
modifies
modified
jur.cit
concerns
earlacts
subsprep
courtdec
earldec
subsdec

VERWIJZ
rechtsgr
cit
wijzig
gewijzgd
cit.jur
betreft
verwijst
verwezen
arresten
eer.jur
lat.jur

RINV
basgiur
atticit
attimod
modsucc
citindir
citindis
attiprec
attisucc
sentenze
giurant
giursucc

KRYDSREF
retsgrl
divcit
aendakt
senaend
citakt
vedrakt
tidlakt
senakt
domme
tidlafg
senafg

MATIERE	SACHGEB	SUB	OND	MAT	EMNE
REPRT	VERZEICH	REGISTER	REPRT	REPRT	REGISTER
DESTINA	ADRESSAT	ADDRESS	GERICHT	DESTINAT	ADRESSAT
LANG.FF	VRB.SPRA	AUT.LANG	AUT.TAAL	LINGUE	AUTSPR
DEPOS	DEPOS	DEPOS	DEPOS	DEPOS	DEPOS
TRA.PREP	VORARB	PREPWORK	VOORBER	LAVPREP	FORBARB
INFO.DIV	DIV.INF	MISC.INF	DIV.INFO	INFODIV	DIVINFO
PARTIES Requer Defend Observ	PARTEIEN Klaeger Beklagt Erklaer	PARTIES Applican Defendan Observ	PARTIJEN Verzoek Verweer Opmerk	PARTI Ricorr Conven Osserv	PARTER Sagsoegr Sagsoegt Indlaeg
NATIONA	NATIONA	NATIONA	NATIONA	NAZIONAL	NAT
TRIB.NAT	VORL.GER	NATCOURT	NAT.RB	GIUDAQUO	NATDOMST
MAGISTR Avocat.g Juge.rap	MAGISTR G.anwal Ber.erst	MAGISTR Adv.gen Judg.rap	MAGISTR Adv.gen Raport	MAGISTR Avvgen Giudrel	MAGISTR Genadv Refdomm
NOTES	ANMERK	NOTES	NOTEN	NOTE	NOTER
PROCEDUR	VERFAHR	PROCEDUR	PROCEDUR	PROCEDUR	SAGSART
LEGISLAT	LEGISLAT	LEGISLAT	LEGISLAT	LEGISLAT	VALGPERI
GROUPE	FRAKTION	POLGROUP	FRACIE	GRUPPO	GRUPPE
dates (sans lexique)	daten	dates (no lexicon)	data	date	datoer



Annex 1B. Updating statistics for CELEX

TABLE 1B.1
 Updating statistics for the analytical part
 (Mbytes transferred from Brussels for Archive
 updating - input in the bases is slightly different because of
 documents automatically generated by programs)

PERIODS	NEW DOCs	UPDATED DOCs	DATE	Mbytes TRANSFERRED
1	95062		1986/05/08	
2	133		1986/12/03	
3	86		1986/12/10	
4	80		1986/12/17	
5	610		1987/01/07	
6	449		1987/01/14	
7	99		1987/01/21	
8	107		1987/01/28	
9	239		1987/02/04	1.36
10	209		1987/02/11	0.42
11	85		1987/02/18	0.48
12	180		1987/02/25	0.98
13	278		1987/03/04	0.86
14	125		1987/03/11	0.51
15	103		1987/03/18	0.37
16	175		1987/03/25	0.47
17	318		1987/04/01	1.13
18	408		1987/04/08	1.49
19	284		1987/04/22	1.02
20	255		1987/04/29	1.06
21	63		1987/05/06	0.32
22	259		1987/05/13	0.98
23	63	469	1987/05/20	0.77
24	161	1010	1987/05/27	0.52
25	84	22	1987/06/01	0.08
26	97	130	1987/06/03	0.49
27	148	528	1987/06/10	0.55
28	354	955	1987/06/17	1.49
29	57	173	1987/06/24	0.20
30	315	37	1987/07/01	1.22
31	88	342	1987/07/08	0.28
32	104	416	1987/07/15	0.20 (?)
33	119	55	1987/07/22	0.45
34	134	110	1987/07/29	0.57
35	178	103	1987/08/05	0.66
36	192	160	1987/08/12	1.38

TABLE 1B.2a
 Texts available on magnetic tape
 (Mbytes/corresponding No of tapes)

month	IC2	PHI	MOS/PER
1/87	-	-	-
2/87	16.83/ 8	-	27.56/ 32
3/87	26.46/10	0.52/6	21.00/ 32
4/87	16.56/ 7	0.43/3	4.70/ 5
5/87	7.69/ 5	0.09/1	3.18/ 3
6/87	2.40/ 1	-	-
7/87	19.47/ 6	-	2.91/ 2
8/87	3.07/ 1	0.26/2	-
SUBTOTALS	92.48/38	1.30/12	59.35/ 61

TOTAL 153.13 Mbytes/111 tapes

TABLE 1B.2b

Text input into CELEX data bases*

month	CLX	CEN	CDE
1/87	-	-	-
2/87	0.59	0.004	-
3/87	-	-	-
4/87	2.12	2.07	2.2
5/87	11.46	0.59	-
6/87	14.41	-	-
7/87	5.05	0.73	0.2
8/87	0.16	-	-
SUBTOTALS	33.79	3.39	2.4

TOTAL 39.58 Mbytes

* some streams (especially SZ) were not working.

TABLE 1B.3a
 Computer resources for a full CELEX updating (Archive-Bases) and extraction of tapes for hosts (week 5-9.10.87)

JCL	Elapsed	CPU	I/O
TITREOP	0.255	0.0261	0.118
CLX010	0.001	0.0000	0.000
CLX020	0.014	0.0008	0.003
CLX030	1.946	0.2039	0.928
EDCELEX	0.035	0.0029	0.013
CLX035	0.629	0.1324	0.444
BN.CLX	0.588	0.0277	0.365
BM.CLX	1.153	0.0533	0.553
BN.CEN	0.545	0.0191	0.253
BM.CEN	1.087	0.0409	0.433
BN.CDE	0.477	0.0163	0.234
BM.CDE	0.885	0.0356	0.387
BN.CNL	0.415	0.0193	0.272
BM.CNL	0.885	0.0409	0.460
BN.CIT	0.824	0.0234	0.460
BM.CIT	1.312	0.0468	0.674
SAVFIL	0.805	0.0058	0.462
Duplicat. remise			
on-line	11.299	0.0152	5.386
Diffus.	1.013	0.0810	0.309
Sav.CLX	1.862	0.0121	0.704
Sav.CEN	5.191	0.0101	0.623
Sav.CDE	1.542	0.0098	0.687
Sav.CNL	1.458	0.0094	0.579
Sav.CIT	0.905	0.0096	0.588
TOTAL	35.126	0.8424	14.935

TABLE 1B.3b
 Disk space occupied by CELEX system
 (week 5-9.10.87) (1 LL = 1.280 bytes)

File	LL	MB
Archive	250.000	320
Espion etc	50.000	64
CEN	564.000	722
CDE	568.000	727
CIT	538.000	688
CNL	519.000	664
CLX	535.000	813
Total	3124.000	3.998

TABLE 1B.3c
Pages printed for CELEX (week 7-11.9.87)

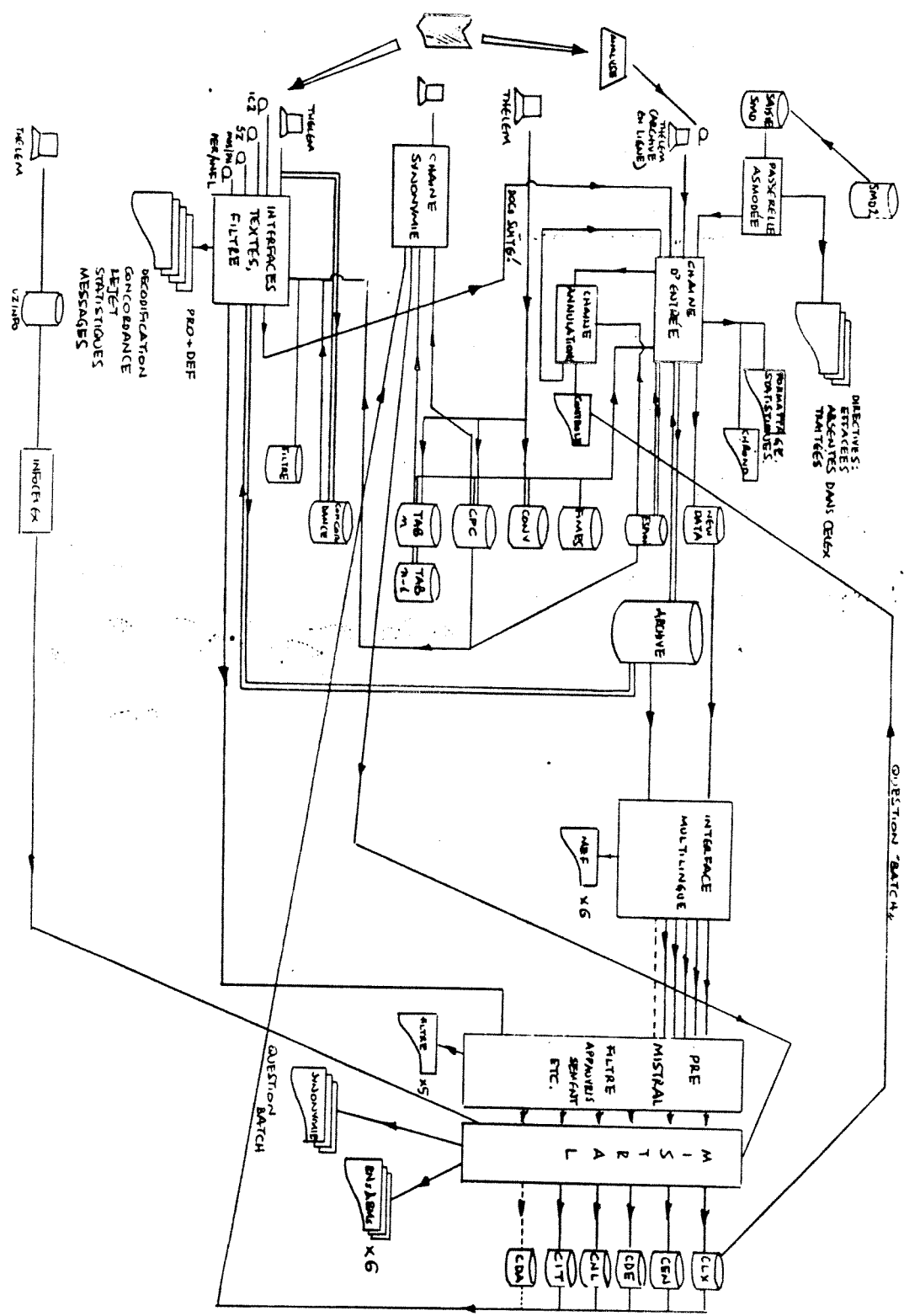
7.9	337
8.9	765
9.9	1162
10.9	2248
11.9	320
Total	4868

Annex 1C. Terminals supported

OLIVETTI TCV450	: 0450
OLIVETTI WS580 &WS581	: 0581
OLIVETTI WS584(VT100 EMUL)	: 0584
QUESTAR DKU 7102 mode SDP	: B702
QUESTAR DKU 7105/107 mode SDP	: B705
VIP7700 or equivalent	: B770
DEC VT52	: D052
DEC VT100	: D100
DEC VT100 (Double density)	: D101
DELTA Model 7300	: D730
GEVEKE Visu 24	: GEVE
ICL - 6402G	: I640
MATRA TTE415	: M415
M I N I T E L	: MTEL
OLIVETTI TCV450	: 0450
OLIVETTI TCV450 (Min.)	: 0451
OLIVETTI WS580 &WS581	: 0581
OLIVETTI WS584 (VT100 emulatio	: 0584
OLIVETTI WS580 &WS581	: 0ZZ1
OLIVETTI WS584 (VT100 emulatio	: 0ZZ4
TELEVIDEO T910 T921	: T910
TELEVIDEO T912	: T912
TELEVIDEO T920	: T920
TELEVIDEO T924	: T924
TEKTRONICS 4107/4109	: TEKT
TELETYPE	: TELE
DEFAULT TYPE	: TT33
Volker - Craig VC 415	: V415
VIP7700 or equivalent	: V770
VIP 78X4 Text Mode	: V784



Annex 1D. The CELEX monster





ANNEX 1E: Chaînes d'entrée CELEX

On peut distinguer trois grandes catégories de chaînes :

- alimentation des données analytiques;
- alimentation des données textuelles;
- chaînes de traitements spéciaux (annulation, synonymie).

Les données analytiques brutes, les bordereaux d'analyse, sont produits par les analystes des équipes CELEX dans les Institutions (Table 2E.1). Ces bordereaux (voir point 2E.a) sont encodés par les pools d'encodage à la Commission (bandes MATRA), la Cour de Justice (bandes PHI) et le Parlement (bandes PHI/RDE). Ensuite, les données sont introduites dans le système (chaîne BIBLIO). Les corrections de la partie analytique des documents déjà présents dans la base se font aussi en ligne (alimentation de la chaîne BIBLIO).

Pour l'alimentation des données concernant les dispositions nationales d'exécution des directives, on se sert de la base ASMODEE via une passerelle (passerelle SMD). Enfin, l'introduction du document INFOCELEX dans les diverses bases se fait via un programme spécial (chaîne INFOCELEX).

Les données textuelles proviennent aussi de sources diverses. Les textes publiés dans les J.O. série L produits par l'imprimerie centrale (80-85 % du volume total) sont disponibles sur bandes (bandes IC2). Le reste, ainsi que les textes anciens qu'il faut rattrapper, sont encodés par des firmes externes (bandes PER, INT, LUX, etc.) selon un plan basé sur le budget disponible.

Les textes publiés dans le Recueil sont envoyés par la Cour de Justice, après leur publication dans toutes les langues, sur les bandes SZ. Des résumés (données textuelles) pour les documents de la Cour de Justice et du Parlement sont aussi disponibles sur bandes (bandes PHI et RPE) similaires aux bandes utilisées pour les données analytiques. Toutes ces bandes sont introduites dans le système via des chaînes spéciales (chaînes TEXTES).

Une dernière catégorie de traitements à l'entrée sert à combler les lacunes du SGBD utilisé (annulation de toutes les références d'un document annulé, - chaîne ANNULATIONS, gestion du thésaurus, - chaîne SYNONYMIE).

TABLE 2E.1 PERSONNEL AFFECTE A CELEX
(hommes/an)

	A/LA	B	C	D	TOTAL
COMMISSION	13	-	2	1	16
PARLEMENT	0,5	3	2	-	5,5
COUR DE JUSTICE	1,5	1,5	-	-	3
CONSEIL	1	-	-	-	1
CES	-	0,5	-	-	0,5
TOTAL	16	5	4	1	26

A ce nombre, il est nécessaire d'ajouter les ressources pour l'encodage, soit 3,5 hommes/an pour la Cour de Justice, 1,8 h/an pour le pool d'encodage de la Commission, et 15085 ECU/an pour encodage externe du Parlement, ainsi que le personnel effectuant des travaux pour l'équipe CELEX de la Commission à savoir :

- 2 hommes/an pour la gestion informatique des bases (Centre de Calcul);
- 1 homme/an pour la maintenance évolutive et corrective (Service de Développement des Applications).

1. CHAINE BIBLIO

La chaîne BIBLIO sert à l'alimentation de la partie analytique des bases. Elle sert à la mise-à-jour de l' ARCHIVE.

La chaîne reçoit en entrée:

- bandes MATRA encodées par le pool d'encodage de la Commission;
- bandes du Parlement européen (partie analytique);
- bandes de la Cour de Justice (les bandes CJ sont formatées sans label : problème pour le CC, à régler);
- plus des corrections de l'Archive en ligne (ARCHIMOD).(Ce système est très rudimentaire et n'autorise pas d'ajout, de contrôle, etc. ... mais presque seulement la substitution de lignes).

Premier traitement : LECTURE (LECTCLX)

- liste par bande : BN
BM - codes : A annulation paragraphe ou ligne
M modification paragraphe ou ligne
R introduction nouveau paragraphe
I insertion MOD.ULT
donne le fichier CLX.MAJ
- liste ARCHIMOD modifications de l'archive en ligne donne le fichier ARCHI.MOD.

Deuxième traitement : CHAINE D'ENTREE FORMATAGE
 CASCADE 3 - 7
 FUSION

A. FORMATAGE (DCXFORM)

Ce programme utilise les fichiers :

- ESPION/DO1 : liste des numdocs/numseq de l'Archive
- CONV : liste des numdocs/numseq Secteur 1
- TAB.CLX : les tables
- FIMES : liste des messages d'erreur
- COPARCHAMP : on utilise : n° des tables de traduction qui doivent être changées + rubriques obligatoires ou non.

1. Création du NUMSEQ à partir du NUMDOC;

2. Contrôle du type de traitement

- non existence de traitement;
- existence des BM.

3. Contrôles de plausibilité

- dates (format, comparaison entre dates ...);
- cohérence interne du document;
 par. ex. : entre FORME et les 3 premiers caractères
 du TITRE, s'il y a un destinataire, s' il y a une
 langue FF, etc...
- vérification sur tables des codes archives.

4. Génération de ULD (par.236) à partir de CAD (par.324)
(i.e. de MOD.ULT à partir de ACT.MOD)

Génération de ULP (par.279) à partir de CAP (par.334)
(i.e. de VIS.ULT à partir de ACT.VIS)

Plus précisément supposons qu'un document A modifie ou vise un autre document B (ou plusieurs documents B1, B2, ...). Pour ce document A sont remplis les champs :ACT.MOD et/ou :ACT.VIS qui contiennent le :NUM.DOC du document B (ou les :NUM.DOCs des documents B1, B2, ...) avec un commentaire sur le type de modification (Table 2E.2). Ce commentaire est structuré en rôle primaire (type de modification de l'acte en son ensemble) et rôle secondaire (type de modification de l'article et/ou annexe de l'acte).

Le programme DCXFORM génère alors pour le document B (ou pour chacun des B1, B2, ...) les champs :MOD.ULT et/ou :VIS.ULT qui contiennent les :NUM.DOCs de tous les documents A modifiant le document B plus un commentaire (l'inverse du rôle primaire et secondaire). De cette façon quand on visualise le document B on voit tous les références des tous les documents qu'ils l'ont modifié, ordonnées en num.doc accendant.

Le formatage ne donne pas lieu à des statistiques

B. CASCADE SECTEUR 3 - 7

Création des docs. secteur 7 à partir des directives.

C. FUSION (DCXFUS2)

1. Mise à jour de l'archive;
2. Donne lieu à la création de deux fichiers :
NEWDATA : données nouvelles en format compatibles
interface (avec code modification);
MINI-ARCHIVE : copie complète de tous les documents créés ou
modifiés dans l'archive.
3. Effectue une deuxième série de contrôles en utilisant
COPARCHAMP
- paragraphes monoligne
- nombre ligne > 99 paragraphes
- 2 docs avec le même numdoc en entrée.
4. Statistiques
5. Edition de CHRONO ARCHIVE (à partir de la mini-archive).
6. Régénère le fichier ESPION.

D. INTERFACE

- Traduit les formes courtes (codes archives) en forme longue en utilisant les tables CELEX (voir table 2E.3).
- Reformate les données (NEWDATA - INPUT MISTRAL).
- Procède, à partir de la MINI-ARCHIVE, au retrait de TOUT champ à modifier et à son remplacement.
En cas d'erreur : message "mot non-trouvé" sur table ****"
(la forme courte erronée apparaît alors telle quelle dans MISTRAL.)

N.B. : Les tables dont le numéro est < 305 sont à même de traduire à partir de codes numériques. Si le numéro est > 305, le numérique est ignoré. (Pour l'interface est numérique toute chaîne de caractères commençant par du numérique.)

Autres messages ex. "double ligne dans le cas d'un paragraphe monoligne"

! se trompe quand on procède au retrait d'un paragraphe.

! les messages interfaces peuvent signaler une erreur ancienne, non seulement donc sur les newdata mais sur toute la mini-archive (très souvent sur ACT.MOD/MOD.ULT).

E. APPAUVRISSMENT

Données BIBLIO

Dans l'archive, seuls les paragraphes titres (TIT.TIC) peuvent d'avoir du semi-riche.

Filter

Some documents in the non-French versions of CELEX do not contain all the fields of corresponding documents in the French CELEX (CLX). (Couverture différenciée) It has been necessary to filter in this way because when these two bases were set up, many documents were no longer valid and it would have proven unnecessarily expensive to load later documents that were no longer useful.

The CELEX-filter consists of the following :

1. Of sector 2, 3 and 4 documents :
 - documents with FIN.VAL >= 1979:07:01 are loaded in full;
 - all other documents are loaded in part (fields NUM.DOC, REFER, REF.PUB, FIN.VAL, MAJ);
2. Of parliamentary documents (sector 5AP, IP, BP, and sector 9) :
 - documents with LEGISLAT = FIRST OR SECOND LEGISLATURE are loaded in full;
 - for all other documents only the fields REFER and MAJ are loaded.
3. All other documents are fully loaded (sectors 1, 6, 7, 5PC, AC, IC, CC).

2. ANNULATIONS

A.ANNULCAD : L'annulation n'est rien, mais la correction est lourde et ne peut se faire que par "insertion cad".

B.ANNULDOC : A pour objet d'annuler un document AVEC TOUTES SES CITATIONS. Il peut y avoir un problème quand le document "père" ne cite pas ses "enfants" (par.ex. le champ CIT.JUR ...)

3. INFOCELEX : Sert à la création des documents INFOCELEX dans toutes les versions linguistiques en y modifiant chaque fois les références de publication (JO, Recueil, etc) des derniers documents entrés dans la base.
4. CHAINES TEXTES : Ces chaînes reçoivent en entrée les bandes textes (IC1, IC2, SZ, PER, etc.).
 - séries PRO : servent à la mise à jour provisoire de l'archive et à la génération des pointeurs dans celle-ci. On évite ainsi l'écrasement des textes déjà existants dans les bases (cas des rectificatifs ou des faux num.docs).
 - séries DEF : appauvrissement plus génération des BM pour introduction dans les bases.

Attention aux documents suite qui se génèrent manuellement.

5. CHAINE SYNONYMIE

Une relation de groupe introduite pour un terme du thésaurus MISTRAL qui ne dispose pas de synonyme s'efface lors de la prochaine introduction d'un synonyme pour ce terme. Le but de la chaîne synonymie est l'introduction pour chaque nouveau terme de certains champs d'un synonyme bidon avant l'introduction éventuelle d'une relation de groupe. Utilise les tables, contrôle les modifications dans les champs concernés et, après détection des différences, crée un synonyme bidon et procède à une MAJ du thésaurus avant la décomposition et l'indexation des termes par MISTRAL.

Rôle primaire	Rôle secondaire	Dispositions affectées	Dates
Rôle primaire caractérisant <u>uniquement</u> une affectation globale			
A abrogation D dérogation E extension (TDC) F confirmation K coordonne P prorogation R remplace S suspension T application différée YD adoption YDP adoption partielle			
Rôle assurant une double fonction (rôle primaire + rôle secondaire)			
E extension I interprétation M modification	E I M	précision obligatoire des dispositions affectées	Date(s) obligatoire(s)
M + rôle secondaire : actes affectés uniquement dans leurs effets particuliers			
M (rôle conventionnel car l'affectation globale ne peut être précisée)	A abrogation ADAP tation D érogation P roroge R emplace S uspension Su ppression T application différée J adjonction C complément	précision obligatoire des dispositions affectées	Date(s) obligatoire(s)
Rôles spécifiques au TDC			
SP RP	SP DD X DD		Date(s) obligatoire(s)

TABLE 1E.3 Type de traitement selon les paragraphes ARCHIVE

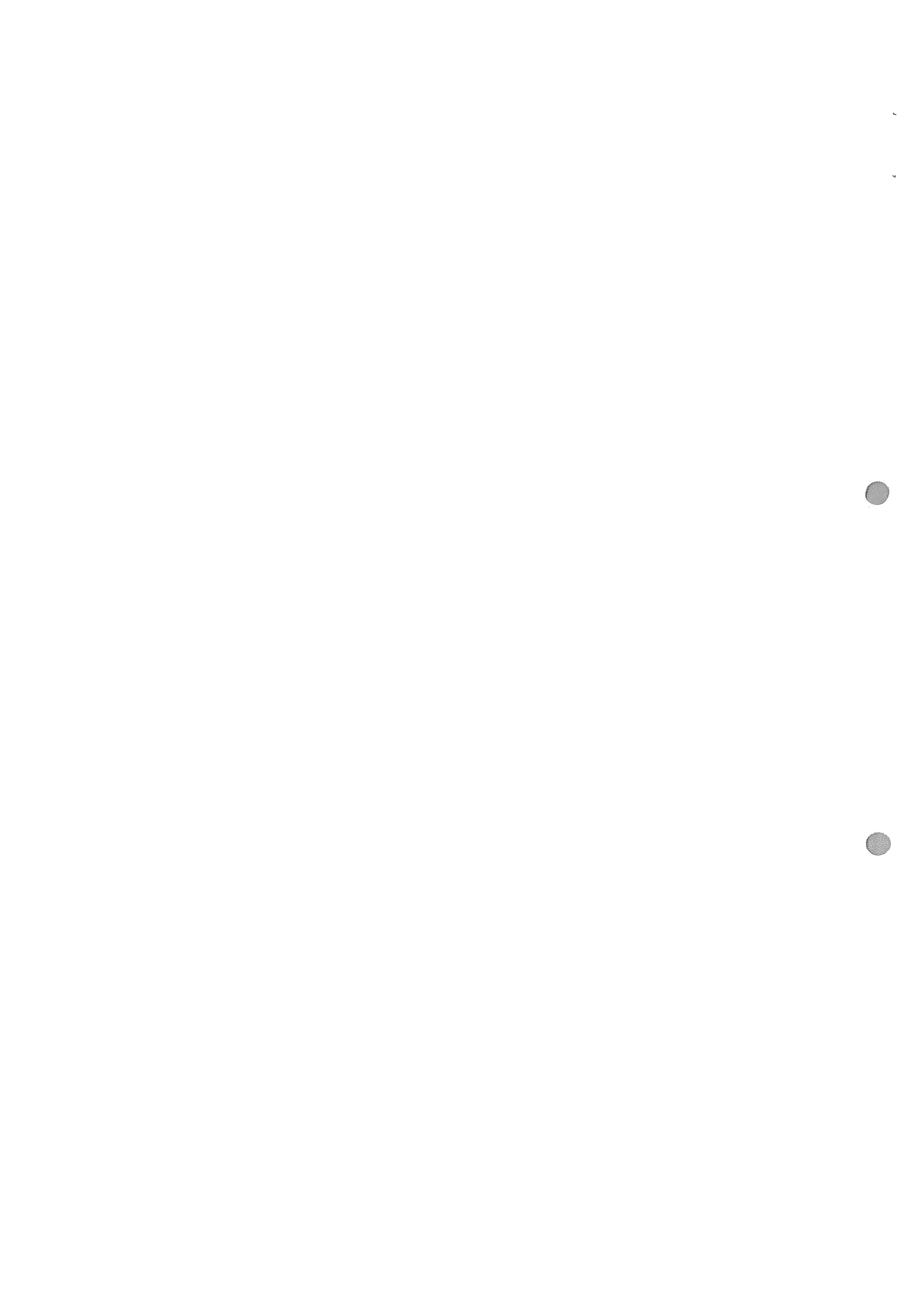
NOM/CHAMP	N° table	§archive	MONO	MULTI	GROP	SYN	PG	INST	Comment.
✓ Act.jur	395	338		X			X	CJUS	
✓ Act.mod	275	324-330 + 334		X			X	COMM CONS CES PE/5 PE/9	rôle primaire gestion passerelle Asmodée (FR) v. Mod.ult/Act.mod/. trad.=14 caractères en une seule chaîne
✓ Act.vis									v. Mod.ult/Act.mod/.
✓ Ant.jur/ ✓ Ult.jur	372	282-302		X				CJUS	- rôle trad. - comment.sans trad; inhibé pr bases non-FR
✓ Arrêts	336	278		X			X	COMM	numérique non trad. trad. à long. fixe
✓ Auteur (a)	010	074	X		X			COMM CONS CES PE/5 PE/9 CJUS	Cascade - 010 - 013 - 014
✓ Auteur (b)	013	074	X		X			PE/5 PE/9	Cascade - 010 - 013 - 014
✓ Auteur (c)	014	074	X		X			CES	Cascade - 010 - 013 - 014
- ✓ Avocat.g/ ge.rap	130	092-093		X				CJUS	
✓ Bas.jur	370	232		X			X	COMM CONS CES PE/5	numérique non trad.
✓ Cit.jur	371	212-24		X			X	CJUS	pour doc. avant 79 comment inhibé pr base non-FR
Dates									Voir: doc./échéance effet/envoi/fin.val/ réponse/transpo.
✓ Débat + dates + dates(v)	345	140-156	X					CES	numérique non trad. comment. accolé à la date (n° session)
✓ Défend	120	090	X			X		CJUS	syn. 1=code min.+ .
✓ Dépos	040	135	X			X		CONS	
✓ Descript	300	410		X		X	X	PE/5	

NOM/CHAMP	N° table	§archive	MONO	MULTI	GROP	SYN	PG	INST	Comment.
✓ Destina	050	082	X		X	X		COMM CES PE/5 PE/9	alimenté par 082 (trad.) + 083 (non trad.)
✓ Doc	365	138						COMM CONS CES PE/5 PE/9 CJUS	sans lexique comment. dates numérique non trad.
✓ Echéance	355	162						COMM	sans lexique comment. dates
✓ Effet (a)	335	152		X				COMM CONS	sans lexique num. non trad. comment. dates
✓ Effet (b)	336	152	X					COMM CONS	num. non trad. comment. accolé à la date
Envoi	340	148						CES PE/5 PE/9	sans lexique num. non trad.
✓ Fin.val	330	158		X				COMM CONS CES PE/5	sans lexique num. non trad. comment. dates
✓ Forme	030	080	X			X		Ttes	contrôle secteur et type, dérivé du num.doc :
✓ Groupe	290	343	X		X	X		PE/9	
✓ Index.cm	301	121	X			X		CONS	
Info.div	400	210		X			X	Ttes sauf PE/9	num. non trad.
✓ Juge.rap									voir avocat.g
✓ Lang.ff	060	116	X			X		COMM CONS CJUS	
✓ Legislat	285	339	X			X		PE/5 PE/9	
✓ Maj									voir mod-ver-maj
✓ Mat ✓ Matières* (a)	070	120	X (Mat)	X (Mats)		X		COMM CONS CES PE/5	sauf secteur 6

NOM/CHAMP	N° table	§archive	MONO	MULTI	GROP	SYN	PG	INST	Comment.
√ Mat √ Matières* (b)	071	120	X (Mat)	X (Mats)		X		CJUS	
√ Mod-Ver- Maj	240	170-172- 174	X					Ttes	intitulé accolé à la date voir table con- stantes 305
√ Mod.ult	270	236à276 + 279		X			X	Ttes sauf CJUS	trad. du rôle pri- maire - table in- terne (code CAD) gestion passer AS- MODEE (FR) voir ci-dessous
√ Mod.ult/ Act.mod/ Vis.ult/ Act.vis	375	236à276 279 324à330 334		X			X	Ttes sauf CJUS	num. non trad. trad. à long. fixe pr rôle secondaire
√ Nationa	140	094	X			X		CJUS	syn 1 = code minéra- logique + .
√ Observ	160	341.	X			X		CJUS	syn 1 = code minéra- logique + .
√ Procedur	100	086	X			X		CJUS	
Ratif	315	146	X	X				CONS	comment. accolé à la date comment. dates
√ Ref.pub	210	069		X				CJUS	secteur 6 trad. à long. fixe
√ Ref.pub	306	047		X				PE/9	num. non trad. trad. à long. fixe
√ Rep	091	128	X					COMM	
√ Repert	090	127	X					COMM CONS	pr contrôle chaîne entrée
Réponse	350	160						PE/9	sans lexique num. non trad. comment. dates.
√ Requer	110	088	X			X		CJUS	syn 1 = code minéra- logique + .
√ Signat	325	154	X	X				CONS	num. non trad. comment. accolé à la date comment. dates

* To be deleted after conversion.

NOM/CHAMP	N° table	§archive	MONO	MULTI	GROP	SYN	PG	INST	Comment.
√ Titre (a)	307	023		X			X	COMM	cascade sect.
√ Titre (b)	316	023		X			X	COMM	cascade sect.
√ Traité	020	078	X			X		Ttes sauf PE/9	
√ Transpos	361	165						COMM	sans lexique comment. date uniq. Direct. + Recomm. CECA
√ Tra.prep (a)	080	124		X				COMM	trad. 1° forme cte de la ligne cascade 080/380
√ Tra.prep (b)	380	124		X				COMM	cascade 080/380 trad. à partir de la 2° forme cte archiv.
Typ/ana	150	126	X					CJUS	
Ult.jur									voir Ant.jur
√ Ver									voir Mod-Ver-Maj
√ Vis.ult									voir Mod.ult/etc.
Constantes	305							Ttes	trad. à long. fixe traitements parti- culiers
√ Disp.nat		185							



Annex 1F. JCLs of CELEX input streams

CLX030A	CHR012 ... CLX035] -	Mise à jour Archive "Analyses"
CLX010	CLX011 (ARCHIMOD)		
CLX040	CLX021		
CLX005	CLX021 (LECTCLX)		
CLX020	CLX021 (LECTCLX)		
CLXNEWD	CLX035 ... CHR012		
ANNO10	annulations		
CLX030B	interface, filtre		
CLX030ND			
MNE010	passerelle MNE		
CLX050B	interface] -	Mise à jour Archive "Analyses PRO"
ARC030	ARC031 ... ARC021		
CLX050A	CHR012 ... CLX035		
MAJ100	Infocalex] -	Mise à jour Bases "Analyses"
MAJ070	pré-traitement		
MAJ070A	pré-traitement		
MAJ080	thésaurus MISTRAL		
COPBMS	"BMS"] -	
MAJBMS2			
MAJBMS3			
MAJBMS			
MAJ090A] -	Biblio et thésaurus
MAJ090B			
JSBMONL	base on-line] -	BN
JSBM	base sur SEMIPERM		
JSBNONL	base on-line] -	BN
JSBN	base sur SEMIPERM		
JSJGONL	base on-line] -	Thésaurus
MAJSG	base sur SEMIPERM		
MAJ060	Bases : anticollision, appauvrissement] -	Mise à jour textes
CLX070	DEF Archive : CHR012 ... CLX033		
CLX060	PRO Archive : CHR012 ... CLX033		

IC2010 IC2030A IC2030B IC2040 IC2070 IC2090 SAVINP	IC2]— —	Chaines Textes		
PER010 PER020A PER020B PER030 PER060A PER060B PER060C	PERFORINCO				
PHI020 PHI030 PHI030VI PHI7000 PHI.REST RESTPHI	PHILIPS				
RPE020	RPE				
SZ010 SZ020 SZ030 SZREST	SZ				
AJUCOUV1 AJUCOUV2	Ajustement de la couverture différenciée				
MAJ070B	chaîne Synonymes				
CXDF41 CXDF42 CXDF43 CXDF44 CXDF51	Diffusion celex				
EDIBIBR EDIPAR	Biblio Paramètres de la base			— —	Editions MISTRAL

ED. CONV
SCRI2429
SCRI3050
SCRI5466
SCRI5574
SCRIS05
SCRIS41A
SCRIS41B
SCRISI05
SCRISI06
SCRISI07
SCRISI08
SCRISI09
SCRISI10
SCRISI11
SCRISI12
SCRISI14
SCRISI15
SCRISI16
SCRISI17
SCRISI18
SCRISI19
SCRISI20
SCRISI21
SCRISI22
SCRISI23
SCRISI51
SCRISI52
SCRISI53

Editions CONV, COPARCHAMP, Tables de traduction

EXTREP
KRONOREP
SAVREP

Extraction Répertoire

ARCHSONL

Archive

CLX035
CLX035B

Espion, Archive et Journal

Mise on-line de
fichiers

CHRONO
STATARCH
SUBARC

Edition / extraction
Statistiques
Substitutions

Divers Archive

CERBERE

Synthèse Cerbère

ALERTE

Systeme d'alerte

4
4



Annex 1G. CELEX operating statistics
(As presented to the Working Party on Legal Data Processing on its
meeting on 11 & 12 June 1987)

A. Within the Community Institutions

The annexes contain statistics on transactions and connection time for 1986 broken down by type of user and language version.

It can be seen from the different charts that the number of users within the various Institutions is increasing constantly for all the language versions.

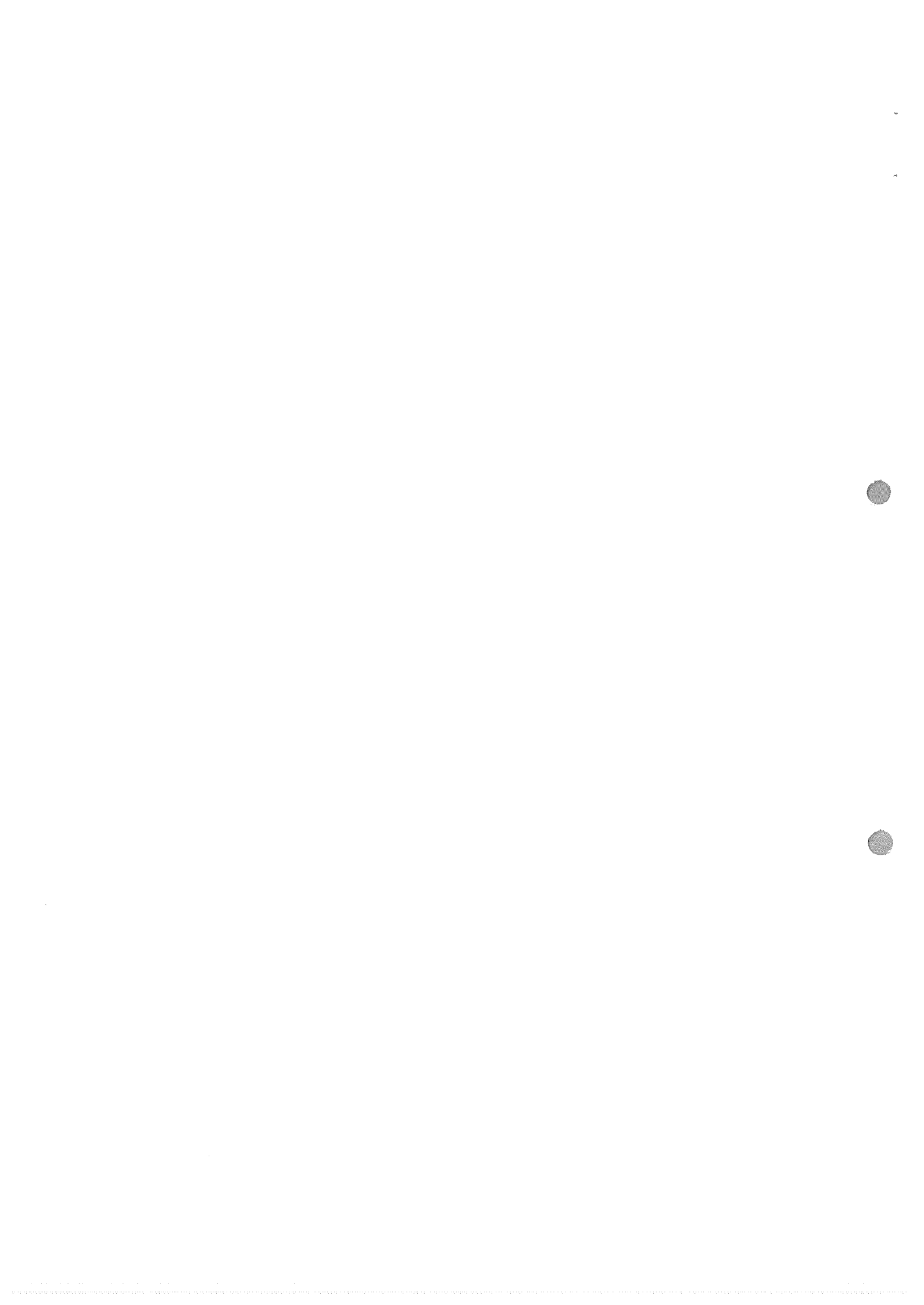


TABLE 1

DISK SPACE
(size in megabytes)

	1982	1983	1984	1985	1986*
CLX	222	277	343	379	812
CEN	0	207	294	334	721
CDE	0	210	314	359	728
CNL	0	0	290	322	664
CIT	0	0	0	325	678
ARCH	166	195	246	246	246

TOTAL:	388	889	1487	1965	3849

situation as at end of each year
* increase due to positional indexation

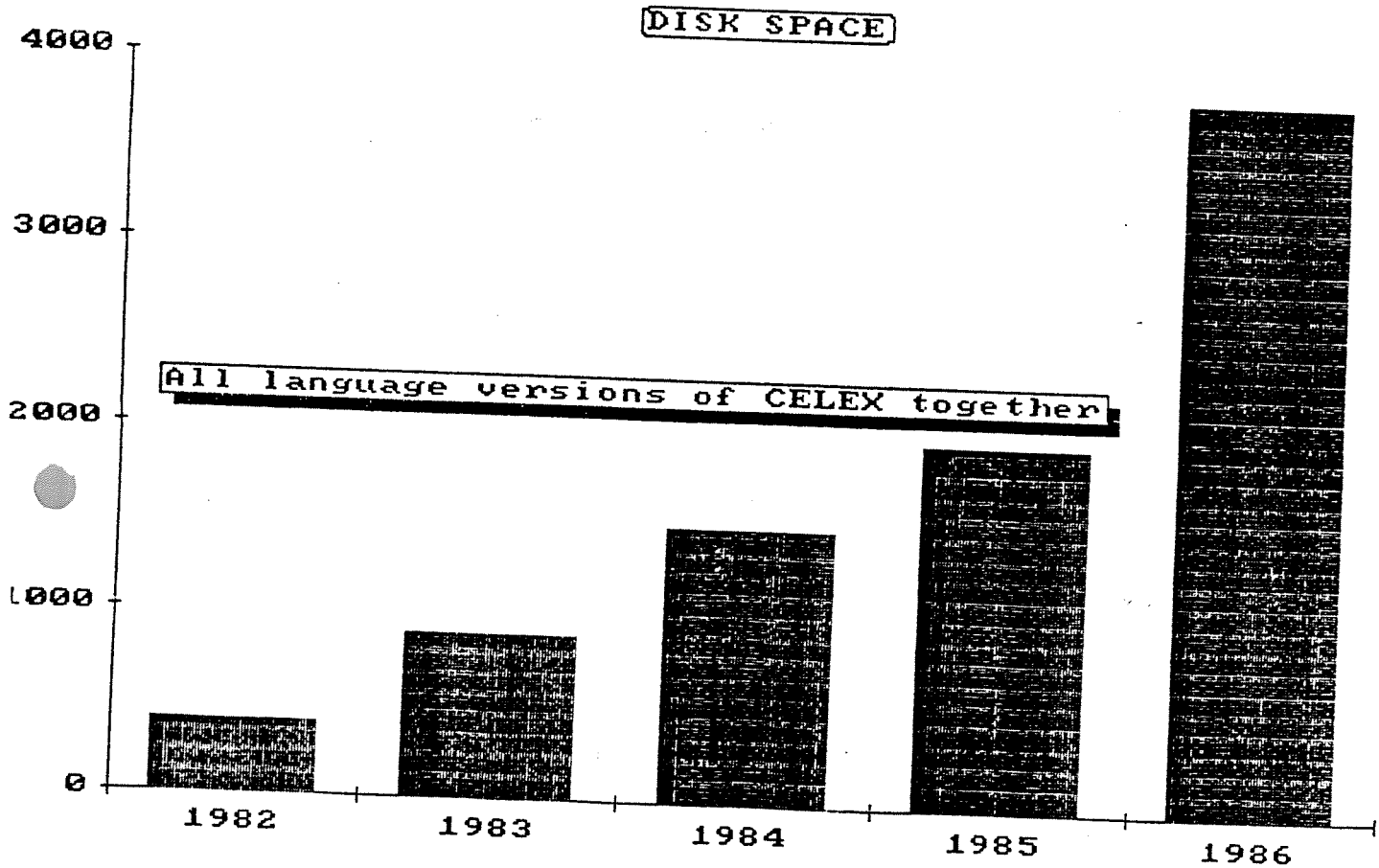


TABLE 2

CONNECTION TIME BY USER CATEGORY (HOURS)

	1982	1983	1984	1985	1986
COMMISSION	2800	3026	5242	7502	9215
OTHER INSTITUTIONS	1380	2031	2391	2881	4070
OUTSIDE USERS	218	439	553	823	2440
TOTAL :	4398	5496	8186	11206	15725

* all language versions of CELEX together

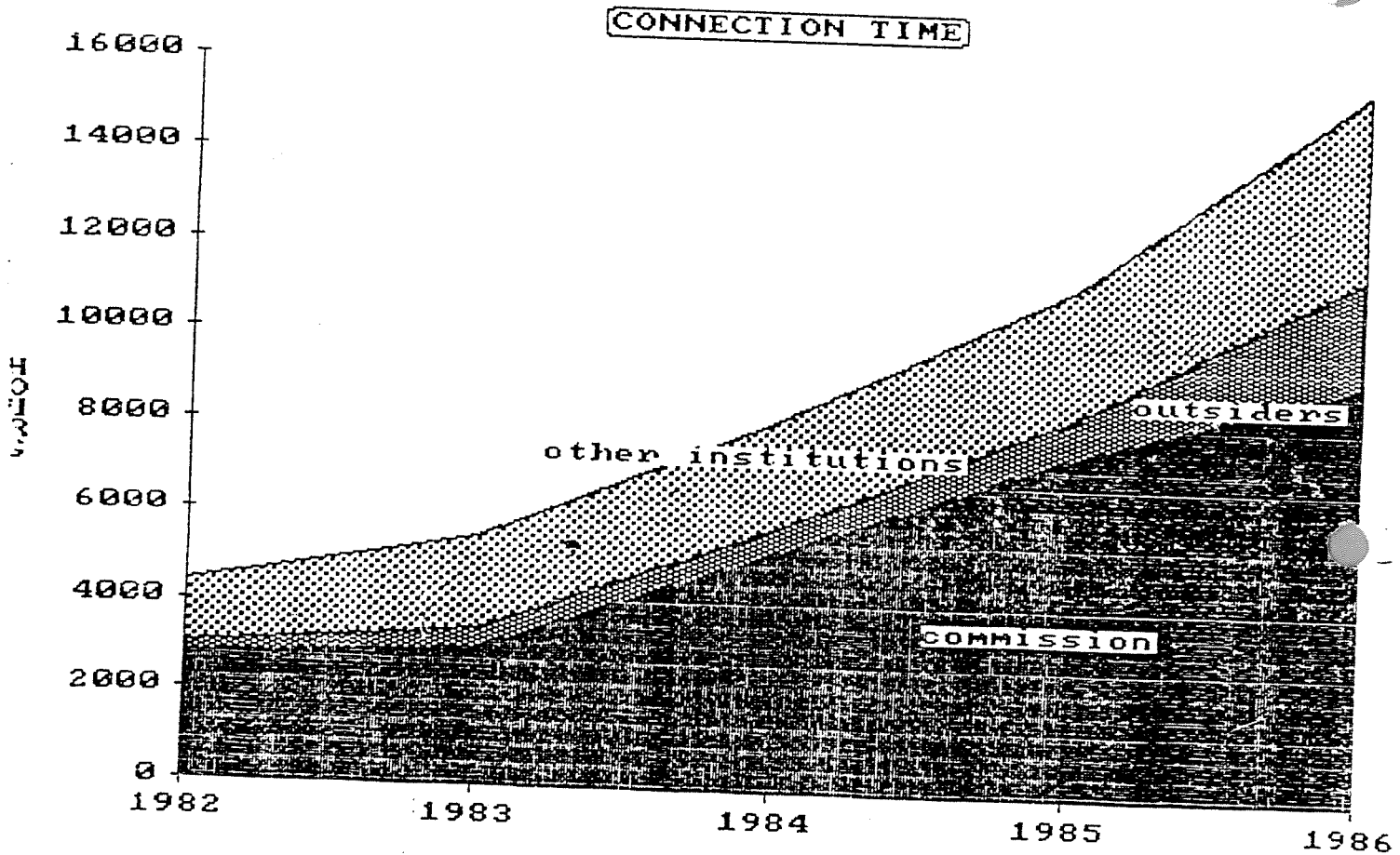


TABLE 3

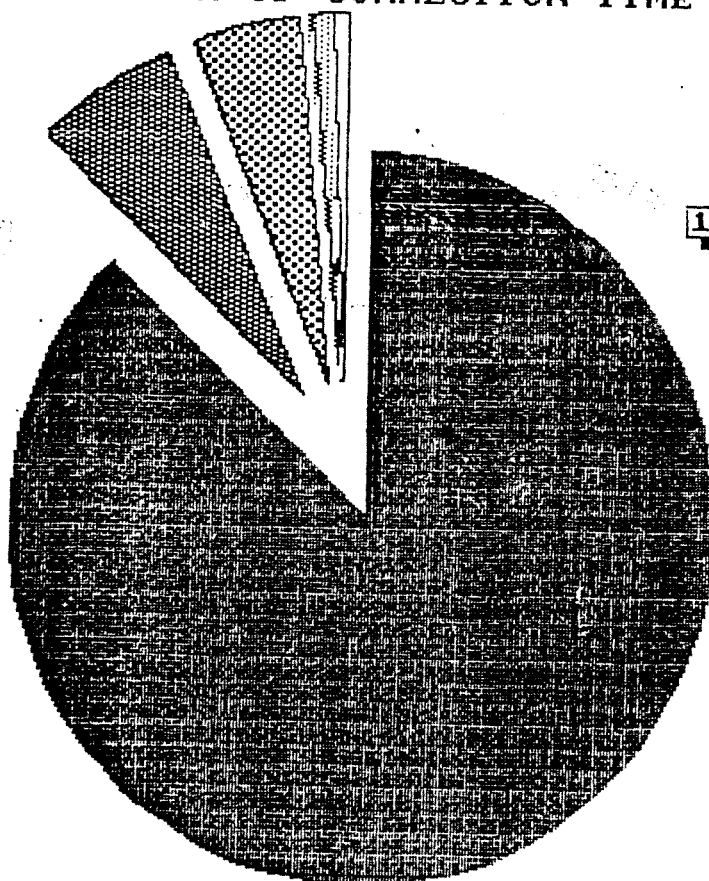
CONNECTION TIME BY LANGUAGE VERSION AND BY INSTITUTION

	CLX			CEN			CDE			CNL	CIT
	1984	1985	1986	1984	1985	1986	1984	1985	1986	1986	1986
EIB	85	43	114	0	0	0	0	0	0	0	0
C/AUDITORS	71	44	42	0	3	1	0	9	3	0	0
ESC	98	84	61	3	3	6	6	1	3	0	0
C/JUSTICE	201	220	785	22	8	6	16	8	3	11	0
COUNCIL	382	795	974	21	30	70	9	17	19	17	0
PARLIAMENT	1335	1452	1765	60	81	108	82	83	60	15	0
COMMISSION	4826	6820	7874	276	485	643	140	197	534	86	77
TOTAL:	6998	9458	11615	382	610	834	253	315	622	130	77

* Figures rounded up to the nearest unit

BREAKDOWN OF CONNECTION TIME

1 = CLX
2 = CEN
3 = CDE
4 = CNL
5 = CIT



■ 1
▨ 2
▩ 3
▧ 4
□ 5

TABLE 4

TRANSACTIONS BY INSTITUTION - ALL LANGUAGE VERSIONS

	1982	1983	1984	1985	1986	% 85/86*
EIB	26715	21897	41961	20177	35210	74.51%
COURT/AUDITOR	19082	13907	15615	13486	10835	-19.66%
CEDEFOP	0	0	0	2150	1948	-9.40%
ESC	31110	33665	57580	50458	34491	-31.64%
COURT/JUSTICE	109436	87525	70689	68690	195257	184.26%
COUNCIL	67702	73958	152137	226891	281248	23.96%
PARLIAMENT	231964	483996	635367	640766	722462	12.75%
SUBTOTAL	486009	714948	973349	1022618	1281451	25.31%
COMMISSION	985656	1065304	1942921	2636961	2798847	6.14%
ALL INSTITUTIONS	1471665	1780252	2916270	3659579	4080298	11.50%

A transaction is taken to mean a carriage return or the transmission of an instruction

* Percentage increase or decrease in transactions from 1985 to 1986

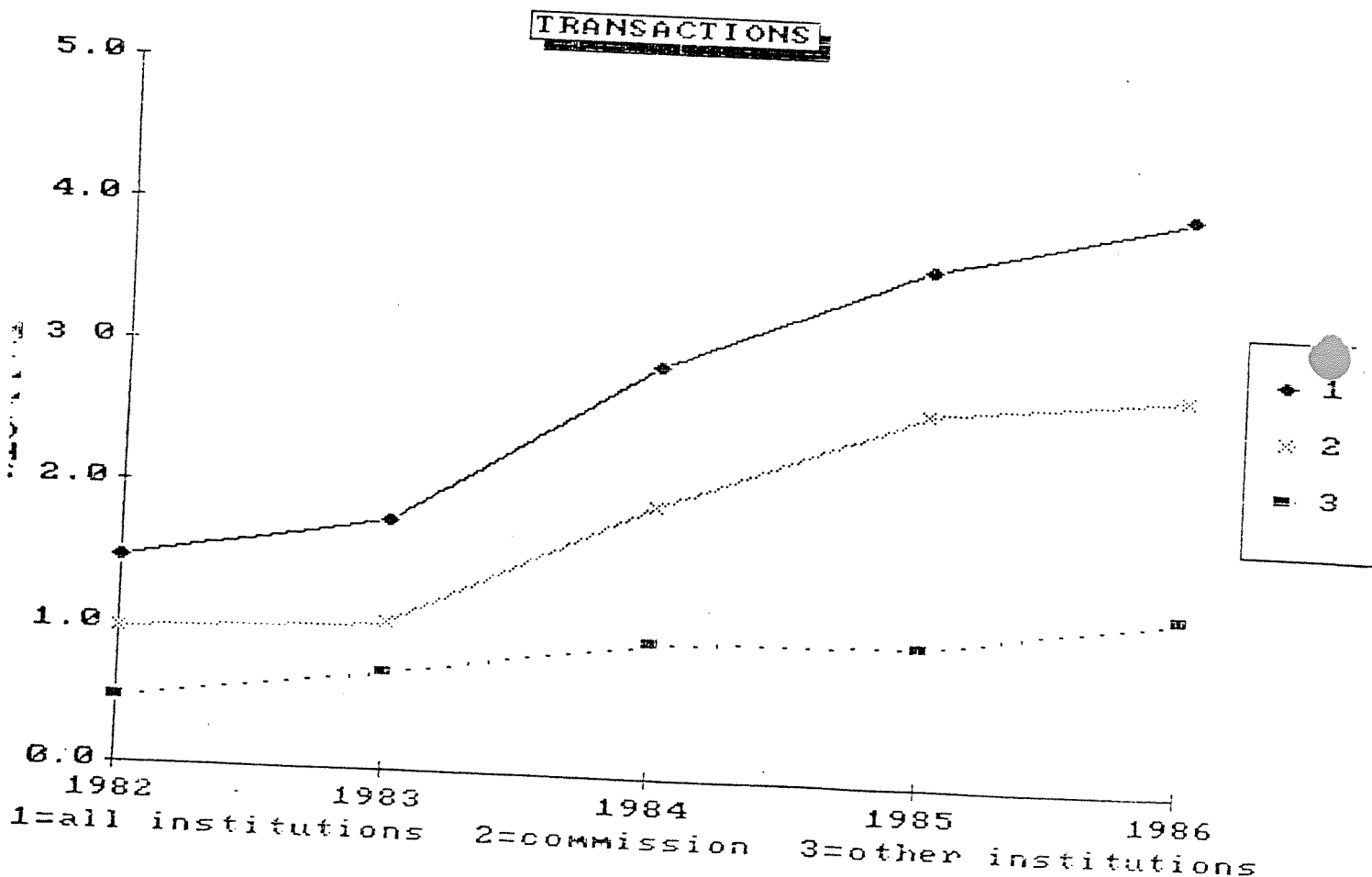


TABLE 5

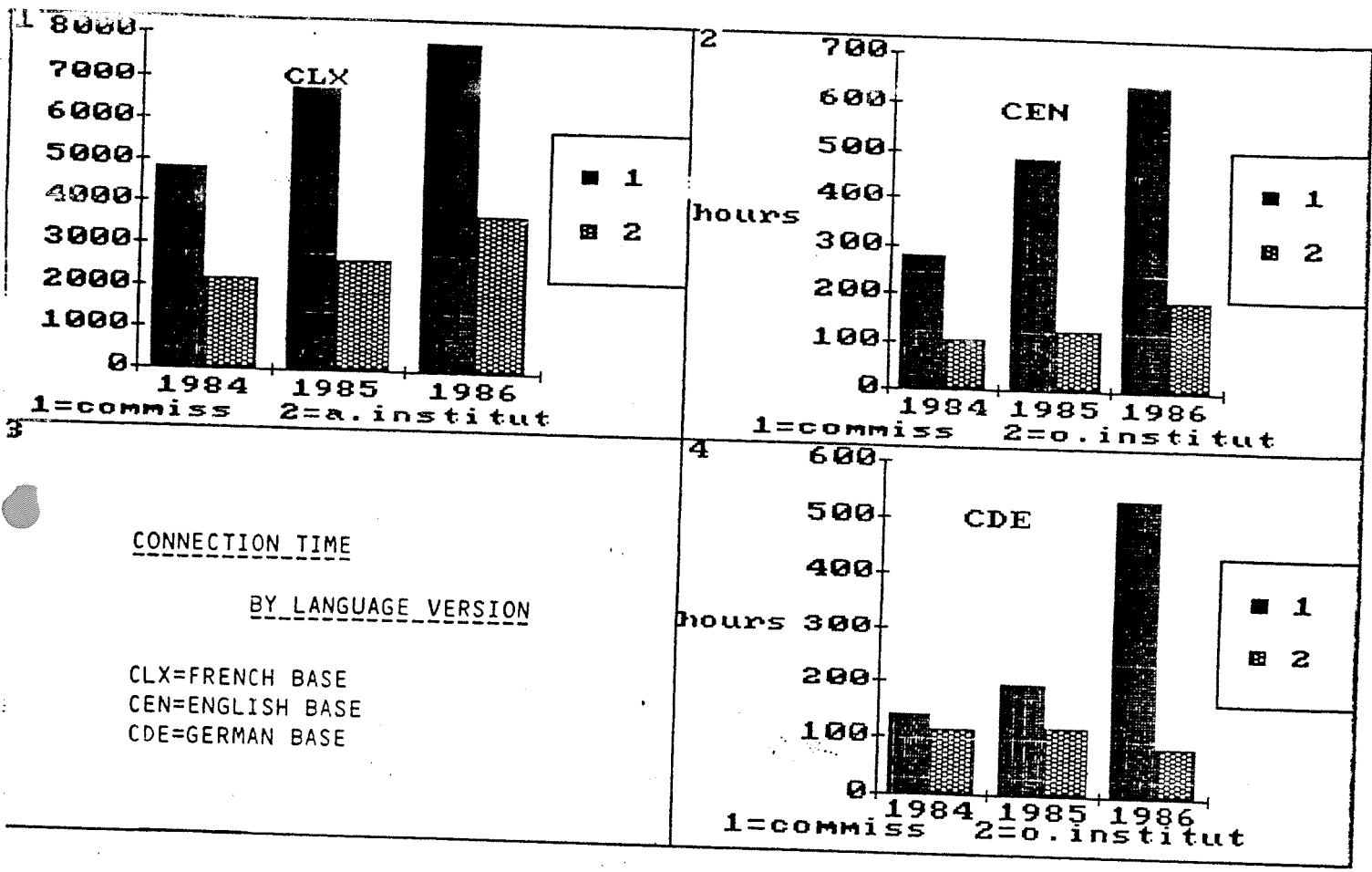


TABLE 6

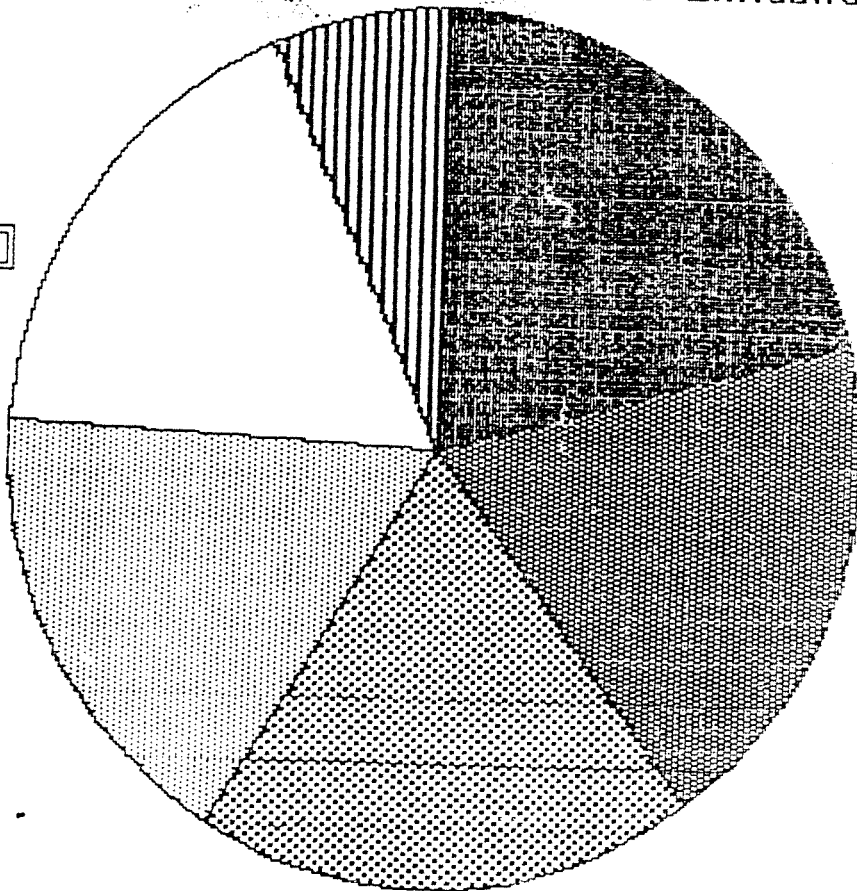
AVERAGE NUMBER OF TRANSACTIONS PER HOUR OF CONNECTION TIME

	1984	1985	1986
EIB	493	469	310
COURT/AUDITORS	219	240	237
ESC	538	573	486
COURT/JUSTICE	295	291	241
COUNCIL	369	269	261
PARLIAMENT	430	396	371
COMMISSION	370	351	302

* A transaction is taken to mean an instruction transmitted (carriage return)

BREAKDOWN OF DISK SPACE BY LANGUAGE

1986



- CLX
- CEN
- CDE
- CNL
- CIT
- ARCH

B. External dissemination

For external users, the Commission provides access to Celex chiefly :

- (i) by offering its customers direct access to its computer centre (on line dissemination) and
- (ii) by disseminating the bases via other bodies (dissemination via hosts).

This new policy has borne fruit in that not only the number of direct-access contracts increased by a third (from 107 to 141) and the connect time by over 50 % (from 823 to 1352 hours) but seven host organizations have already signed redistribution contracts. Three further contracts are imminent.

The statistics on on-line dissemination (see part A) show that :

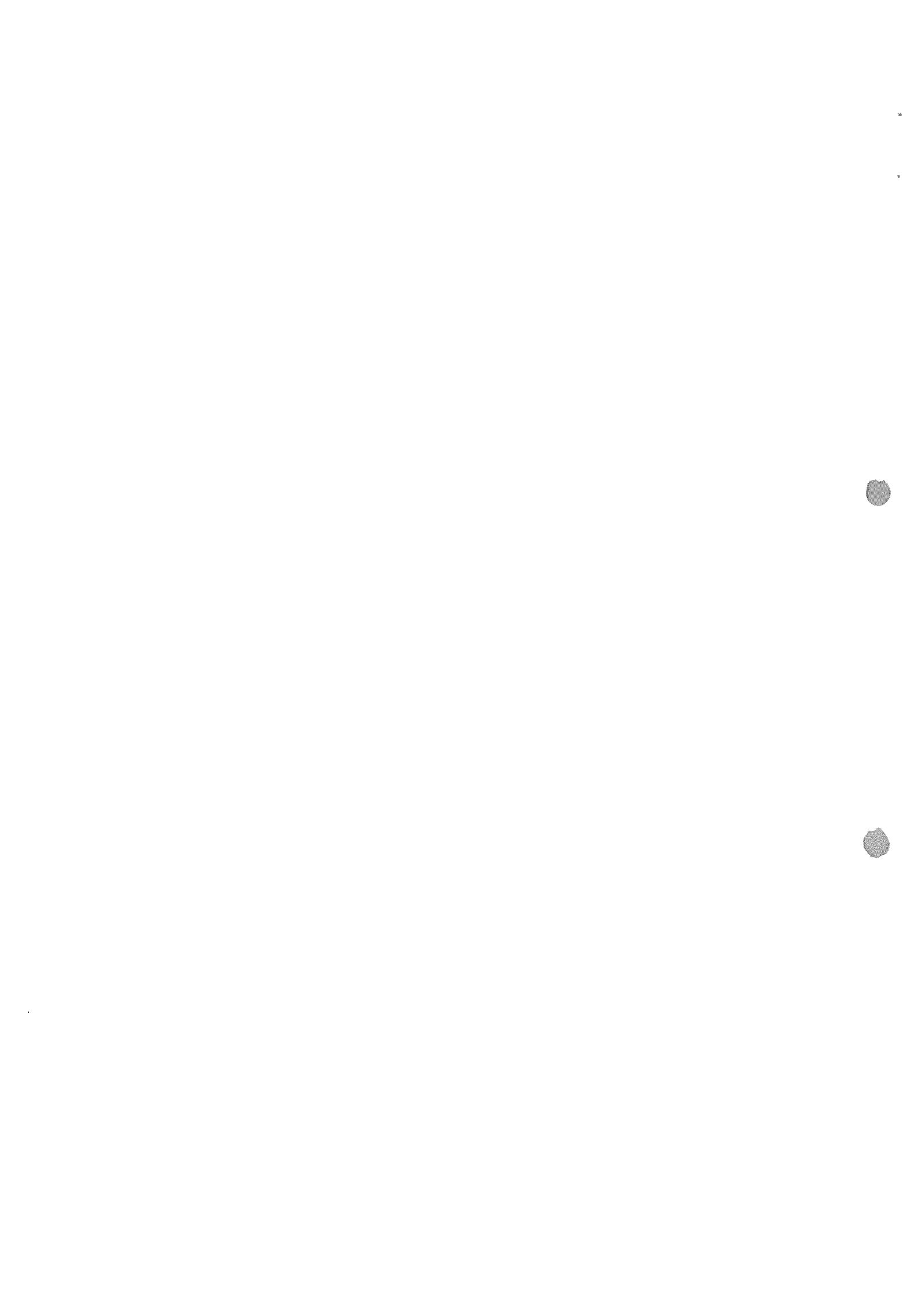
- (i) privileged users (universities, European documentation centres, permanent representatives) account for only one third of users (table 3);
- (ii) the average connect time is one hour per month (11.49 hours per year - see tables 4 and 5);
- (iii) by far the most use is made of the English version of Celex (see table 6).

Of the seven hosts that have already signed contracts (see part B) :

- (i) only one (NOMOS) has opted to run the complete base from its own operating centre;
- (ii) three (DATALEX, CNIJ, CONTEXT) operate only part of the base from their own centres;
- (iii) one (INFE) offers its customers gateway access via its operating centre;
- (iv) two (TIME-SHARING, CREDOC) have concluded distribution-agent contracts.



Annex 2. DATA ON THE NEW CELEX



Annex 2A. Additional full text coverage

Table 2A.1: Millions of characters to be added by sector and language to obtain the new CELEX (calculations based on an average document length = 8000 characters)

SECTOR	CLX	CEN	CDE	CNL	CIT	CDA	CEL	CES	CPO	RESPONS.
1	0.0	0.0	0.0	5.0	5.0	5.0	5.0	5.0	5.0	
2+	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	CONN.
3a	36.5	71.9	71.9	128.6	154.3	154.3	154.3	95.0	95.0	CONN.
4	0.8	1.3	1.3	2.0	2.0	2.0	2.0	1.4	1.4	CONN.
5§	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	CONN.
	0.8	0.4	2.8	1.7	1.8	1.8	1.8	1.8	1.8	PARLW.
6!	17.0	17.0	17.0	17.0	31.0	31.0	31.0	31.0	31.0	CJUS
7%	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	CONN.
8	?	?	?	?	?	?	?	?	?	
9*	70.7	64.5	64.5	64.5	64.5	64.5	60.3	8.2	8.2	PARL/CONN
10	?	?	?	?	?	?	?	?	?	
SUB-TOTALS	149.8	179.1	181.5	242.8	282.6	282.6	278.4	166.4	166.4	
TOTAL	1929.6									

Remarks:

+ for Sector 2 only documents not covered in Sector 3 and without text are considered; estimate of some 50 valid decisions of Joint Bodies adopted.

@ Sector 3 contains adopted Sector 2 documents.

\$ For Sector 5 only Commission proposals pending before the Council are considered; for Parliamentary documents of the IP type a brief summary is considered.

% Sector 7 will contain the basic Treaties (some 1,2 Mcharacters) and the consolidated versions of valid documents with amendments (as at 9 Sept 1987 there were 1664 such documents for Sectors 3 and 4 i.e. 12,85 Mcharacters).

* only parliamentary questions of 1st and 2nd Legislature were considered; for EL, ES, PO only questions after the respective accessions were considered; for each document the calculation was based on 1/2 page of question and 3/4 page of reply.

! Sector 6 estimate does not consider "recours" and cases not accepted.

Table 2A.2: Texts available on magnetic media (IC2, FORMEX, SZ, others) (Mcharacters).

SECTOR	CLX	CEN	CDE	CNL	CIT	CDA	CEL	GES	CPO
1	0.0	0.0	0.0	5.0	5.0	5.0	5.0	5.0	5.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3+	0.0 20.1	0.0 24.1	0.0 24.1	0.7 24.1	0.8 24.1	0.8 24.1	0.8 24.1	0.0 95.0	0.0 PRE-FORMEX 95.0 FORMEX
4	0.3	0.4	0.4	0.4	0.4	0.4	0.4	1.4	1.4 FORMEX
5	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3 FORMEX
6a	17.0	17.0	17.0	17.0	31.0	31.0	31.0	31.0	31.0
7%	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5
8	?	?	?	?	?	?	?	?	?
9*	70.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	?	?	?	?	?	?	?	?	?
SUB-TOTALS	123.9	57.3	57.3	63.0	77.1	77.1	77.1	148.2	148.2
TOTAL AVAILABLE	829.2								
DIFFERENCE TO BE ENCODED	25.9	121.8	124.2	179.8	205.5	205.5	201.3	18.2	18.2
TOTAL TO BE ENCODED	1100.4								
OPOCE COMMITMENT	40.0	45.6	42.4	52.0	52.0	52.0	130.0	0.0	0.0
TOTAL COMMITMENT	414.0								
TO BE PROVIDED	-14.1	76.2	81.8	127.8	153.5	153.5	71.3	18.2	18.2
TOTAL TO BE PROVIDED	686.4								

Remarks:

+ assuming that 80% of pre-FORMEX texts are available on IC2 tapes for CNL, CIT, CDA, CEL.

@ Sector 6 documents are to be provided by the Court of Justice.

* for CLX via EC01.

% assuming that consolidated versions of Treaties are available on FORMEX tapes and consolidated versions of Sector 3 and 4 documents are produced by downloading/uploading.

Annex 2B. Extra analysis time and volume of input data to produce a base as described in paragraph III.1.

SECTOR	ARCHIVE DOCS	BACKLOG ANALYSIS h/an	ADDITIONAL ANALYSIS UNDER NORMAL CONDITIONS h/an	RESPONS
1		0.250	0.000	COMM
2				COMM
3 a	173	0.024	0.006	COMM
b	2000	0.020	0.040	
c	?			
4		0.000	0.000	COMM
5	124	0.450	0.360	PARLM
6	450	0.090	0.060	CJUS
7	1164	4.050	1.440	COMM
8				
9				PARLM
TOTAL	3911	4.884	1.906	

Remarks:

- a. Decisions of Joint bodies
- b. Erase links with proposals adopted
- c. National Measures Implementing Directives



Annex 2C. Names and contents of fields and their grouping into inverted files (lexica) for the language versions of the new CELEX.

The new CELEX should comprise six inverted file groups (lexica) with the following field breakdown and description (names for lexica/fields are given for the French version of CELEX; corresponding names for the other language versions will follow later.

1. BIB equivalent to the existing :BIBLIO but not implicit any more.
 - a. NUMERO (alphanumerical, keywords, all documents/Sectors) containing the standardized identification number of the document (:num.doc) as it exists today but with some adjustments as far as some Sectors are concerned. This field should have a synonym name ND (of CCL) or ID (of SCAD) or whatever the general name of the identification number's field will be in an harmonized data base environment. The information contained in the identification number does in fact duplicate information found in other fields; however, this number permits an appropriate linkage between documents, is too old a habit for CELEX users to change, and permits a sorting of CELEX documents in an appropriate way.

The identification number of CELEX documents will be standardized as follows:

position 1 : Sector number (for Sector 10 documents the number 0).

positions 2,3 : year of document attributed according to existing rules.

position 4 : document type; for Sector 5 documents a single code is to be given to conform to other Sectors. For the transcription of the two-letter code now in use to one-letter codes as proposed above, two solutions can be envisaged; the adoption of the correspondence that now exists for the NUM.SEQ (:refer) construction i.e. PC=A, AP=B, IP=C, BP=D, AC=E, IC=F, AJ=G, CC=H, AG=J, AK=K, DC=I; or the adoption of the following more or less mnemonic correspondence: PC=P, AP=C, IP=I, BP=B, AC=S, IC=T, AJ=J, CC=M, AG=G, AK=K, DC=D.

positions 5 to 8 and, where appropriate, 9 to 12: the number of the document given as follows:

* Sector 1 : each article of the treaties is considered a separate documentary unit and is given a 4-digit number; e.g. Article 85 will be given the number 0085. Sub-articles (i.e. bis, ter etc.) are given corresponding serial numbers in brackets that occupy positions 9 to 12 e.g. Articles 78 and 78bis of the ECSC treaty will have the numbers 151K0078(01), 151K0078(02) respectively. Annexes, declarations, protocols etc. are given "ad hoc" 4-digit numbers where the first position could be a letter indicating the nature of the attachment to the treaty (A=annex, P=protocol, D=declaration etc.) and the following three positions are occupied by corresponding numbers; attention should be given to the proper sequence of these documents within the treaty and it must be stipulated that annexes always come before declarations which come before protocols. If that is not the case, then letters or numbers should be given arbitrarily and in a way that will reproduce the original sequence of the attachments in the treaty. Numbers within brackets can also be given in the case of subarticles in the attachments.

* Sector 2 : a 4-digit number is formed using the date of signing or publication; serial numbers in brackets are used for documents signed or published on the same date.

* Sectors 3, 4 : the 4-digit number of the document is used; documents without a number receive one 4-digit number formed using the date of publication. Double numbers are avoided by attaching a serial number in brackets.

* Sector 5 : for Commission proposals the final COM number is used; for COM documents generating more than one documentary unit, serial numbers in brackets are added. For European Parliament resolutions the identification number is the number of the document of the part-session. For ESC opinions an internal number is used.

* Sector 6 : the number of the case is used; occasionally a serial number in brackets is also used, e.g. for several decisions concerning the same case.

* Sector 7 : the number of the original document is used.

* Sector 8 : (p.m.)

* Sector 9 : for written and oral questions asked during question time, the original number is used; non-question time oral questions are given an internal number.

* Sector 10: (p.m.).

There are overflow documents in the CELEX system, as there are in all full-text databases under documentary DBMSs, and therefore the extension to the identification number (a dot and a serial number) will have to be kept.

b. TYPE (alphanumeric, keywords, all documents/Sectors) containing the type of document i.e. information now present in the fields :typ.doc and :forme, and part of the information in the field :index.cm (if information is present). Information should be given in the documents in "plain-words" but it should also be possible to search with an abbreviated form, using synonymy. A special code "suite" should be envisaged for overflow documents in the base. Field synonym name DT (of CCL).

c. AUTEUR (alphanumeric, keywords, all Sectors) containing the name of the author (person, Institution) of the document i.e. information at present in the fields :auteur and :legislat and part of the information in the fields :groupe (for Parliamentary Questions) and :avoc.g (for opinions of the Advocates-General). Field synonym name AU (of CCL).

d. REF.PUB (alphanumeric, text, all Sectors) containing the publication reference of the document i.e. information at present in the field :ref.pub. The contents of this field will be standardized to replace the existing field :pub. The date of publication must be formatted as in date fields (YYYY/MM/DD), in order to enable it to be used instead of the present field :pub in, for example QI procedures. OJ numbers should then be formatted as one character string (i.e. C-198, L-197) while page numbers should be given 5 digits (i.e. 01985, 01987). Field synonym name SO (of CCL).

e. DESCRIPT (alphanumeric, keywords, all Sectors) containing keyword classification of the document i.e. information at present in the fields :descript and :mat. The short term objective is to use a standardized thesaurus to fill this field for all the documents in the base. Field synonym name CT (of CCL) or CL (of SCAD).

f. REQUER (alphanumeric, keywords, Sector 6) containing the name(s) or type(s) of the applicant (person, Institution) in a case before the Court i.e. information at present in the field :requer.

g. DEFEND (alphanumeric, keywords, Sector 6) containing the name(s) or type(s) of the defendant i.e. information at present in the field :defend.

h. OBSERV (alphanumeric, keywords, Sector 6) containing the name(s) or type(s) of observers i.e. information at present in the field :observ.

i. INTERV (alphanumeric, keywords, Sector 6) containing the name(s) or type(s) of the intervenor(s).

j. NATIONA (alphanumeric, keywords, Sectors 6, 9) containing the nationality of the applicant, the defendant, the national court, or the MEP, i.e. information at present in the field :nationa and part of the information present in the field :groupe.

k. TRIB.NAT (alphanumeric, text, Sector 6) containing references to proceedings before the national court(s) in the context of a preliminary ruling i.e. information at present in the field :trib.nat.

l. AVOCAT.G (alphanumeric, ^{keywords} text, Sector 6) containing the name of the Advocate-General of the case i.e. information at present in the field :avoc.g. This field should not be filled in for the opinions of Advocates-General because of the obvious redundancy (see field AUTEUR).

m. JUGE.RAP (alphanumeric, ^{keyword s} text, Sector 6) containing the name of the Judge-Rapporteur of the case i.e. information at present in the field :juge.rap.

n. PROCEDUR (alphanumeric, keywords, Sector 6) containing the nature of the proceedings i.e. information at present in the field :procedur.

o. REPERT (alphanumeric, keywords, Sectors 1-7) containing a classification of acts according to the codes of the Directory of Community legislation in force, i.e. information at present in the field :repert. In this same field it will be possible to interrogate using not only the numerical classification code of the Directory but, through synonymy, using the "plain words" equivalent or the abbreviated form now used in the field :mat. In fact the new REPERT field should contain the information already present in the present field :repert plus its "translation".

p. LANGUE (alphanumeric, keywords, Sectors 1-4, 6-8, 10) containing the language of the proceedings i.e. information at present in the field :langue.ff. Field synonym name LA (of CCL).

q. DESTINA (alphanumeric, keywords, Sectors 2-5) containing the name(s) of person(s)/Institution(s) to which the act is addressed i.e. information at present in the field :destina.

r. DEPOS (alphanumeric, keywords, Sector 2) containing the depository body for agreements i.e. information at present in the field :depos.

s. 12 DISP.XX fields (alphanumeric, text, Sector 3L) where XX will be the country code i.e. FR, UK, NL, DE, DA, IT, ES, PO, EL, BE, LU, IR. These fields will contain the relevant national provisions implementing directives i.e. information at present in the field :disp.nat, structured as follows:

DISP.FR: (corresponding publication reference for French provision(s) implementing corresponding directive).

DISP.UK: (corresponding publication reference for United Kingdom provision(s) implementing corresponding directive).

DISP.ES: not yet implemented. etc. etc.

t. INFO.DIV (alphanumeric, keywords, Sectors 2-7) containing miscellaneous information at present in field :info.div and partly in fields :ref.int, :comment and :index.cm.

u. RESP (alphanumeric, keywords, Sectors 3-4) containing the department responsible for the document i.e. information at present in the field :resp. This field should be kept secret from external users.

v. ASSO (alphanumeric, keywords, Sectors 3-4) containing the associated department i.e. information at present in the field :asso. This field should be kept secret from external users.

2. REL equivalent to the existing :rel lexicon. In order to take full advantage of the Selection Deduite feature of MISTRAL V5 that corresponds to similar features in other DBMSs (expand, zoom, etc.), some of the fields of the lexicon have to be of the "keyword" type (monoseparator) while others must contain the identification number dissociated from other characters. However, the fields of the present :rel lexicon contain further information, such as the nature of amendments, which must be kept at least in part. After considerable thought it is proposed that fields of the new REL lexicon be devised as follows:

a. BASE (alphanumeric, text, Sectors 2-7) containing the identification numbers of documents forming the legal base of the act i.e. information at present in the fields :bas.jur and :traite. Data now present in the field :traite will be covered by the use of synonymy

between the identification number and the full name of the treaty.
Information will be structured as follows:

identification number article-n,m,l alinea-k,i ;#

b. CIT (alphanumerical, text, all Sectors) containing identification numbers of acts cited in the document i.e. information at present in the fields :cit.div and :cit.jur. Information will be structured as follows:

id.number article-n,m,l alinea-i,k N-i,j,.. P-o,t,.. ;#

where N, P have the same significance as they have in the present :cit.jur field.

c. MOD (alphanumerical, text with positional indexation, all Sectors) containing the acts amended by the document i.e. information at present in the fields :act.mod, :act.vis, :act.jur and :art.jur according to the following format:

id.number type-of-modification-doc. from-870101 till-871231;# or

id.number type-of-modification-art. n,m,k from-870101

till-871231 ;# or

id.number type-of-modification-ann. i,j,o from-871231

till-890101 ;#

For amendments of the same type to different articles of the same document (i.e. same id.number) the number is given only once. Data are ordered according to ascending id.number.

d. ULT (alphanumerical, text, all Sectors) containing the acts that have amended the document i.e. information at present in the fields :mod.ult, :vis.ult and :arrets, according to the following format:

DOCUMENT type-of-modification id.number from-.. till-.. ;# or

ARTICLE n,m, type-of-modification id.number from-.. till-.. ;# or

ANNEX k,l, type-of-modification id.number from-.. till-.. ;#

For more than one amendments of the same article by the same document the article and its number are given only once. Data are ordered according to ascending article number.

e. TRA.PREP (alphanumerical, text, all Sectors) containing information on the preparatory work on the document at present in the field :tra.prep, according to the following format:

id.number (if it exists) structured comment.

f. NOTES (alphanumerical, text, Sector 6) containing information at present in the field :notes.

3. RELAT; this special inverted file grouping (lexicon) is devised in order to contain two (monoseparator) fields that cannot co-exist with the other (positionally indexed) fields of the REL lexicon in the context of MISTRAL V5. It will be kept secret from the users. Selection Deduite questions could then operate on the lexicon by means of a special macro-procedure accessible to all users.

a. BAS.1 (alphanumerical, keywords, Sectors 2-7) containing the identification numbers of documents forming the legal base of the act i.e. information at present in the fields :bas.jur and :traite.

c. CIT.1 (alphanumerical, keywords, all Sectors) containing identification numbers of acts cited in the document i.e. information at present in the fields :cit.div and :cit.jur.

4. DATE lexicon corresponding to the present :date lexicon with the following field breakdown:

a. DOC (date, all Sectors) containing the date of the document, at present in the field :doc.

b. DEM (date, Sectors 6, 8) containing the date the application was lodged, at present in the field :dem.

c. EFFET (date with characteristics, Sectors 1-5) containing the date an act enters into force, at present in the fields :effet and :notif according to the following format:
YYYY/MM/DD ENTREE-EN-VIGUEUR; YYYY/MM/DD IMPLEMENTATION; YYYY/MM/DD NOTIFICATION;

d. FIN.VAL (date, Sectors 1-5) containing the date of end of validity, at present in the field :fin.val.

e. TRANSPOS (date, Sector 3) containing the date by which the Member States must comply with a directive, at present in the field :transpos.

f. SESSION (date with characteristics, Sectors 5,9) containing the dates of the corresponding Session or part-Session and voting of the E.P. i.e. information at present in the fields :debat and :vote according to the following format:
YYYY/MM/DD DEBAT ; YYYY/MM/DD VOTE ;

5. TXT lexicon corresponding to the present :texte and :recueil lexica with the following field breakdown (all fields will be alphanumerical, positionally indexed):

Sectors 1-5, 7 with full text

Sector 6

TITRE (alphanumerical, text, positional indexation, all documents/Sectors) containing the title of the document, at present in the fields :titre and :tit.jur. Field synonym name TI of CCL.

VISA (with 5 overflow fields)	SOMMAIRE
CONSID.(with 5 overflow fields)	INDEX
CORPUS (with 5 overflow fields)	INTITULE (= :intit.1,
FOOTNOTE	:intit.2, :cour)
ANNEXE (with 10 overflow fields)	MOTIFS (with 10 overflow
RECTIF 1-5	fields)
	VISA (for older docs)
	DEPENS
	DISPOSI
	CONCLUS (with 10 overflow
	fields)

For Sector 5 and (possibly) 9 documents, there would be the field RESUME, with an abstract of the document (i.e. corresponding to the existing field :resume). Field synonym name to be provided for this field is AB (of CCL).

6. GES lexicon corresponding to the existing :ges lexicon. The most important information concerning the updating of a document is of course:

- (i) how many times a document has been updated,
- (ii) how many times a field has been updated.

Unfortunately MISTRAL V5 has no answer to the above-mentioned problems and the development of such tools in the framework of the CELEX development plan cannot be envisaged.

The GES lexicon will contain the following fields (all the fields of this lexicon should be under secret for external and internal users being visible only for the members of the CELEX teams in the Institutions):

a. REP (alphanumerical, keywords, Sectors 3,4) containing information at present in the field :rep (usefull in producing the Directory).

b. ANALYSTE (alphanumerical, text, all Sectors) containing the name of the analyst of the document (i.e. information at present in the field :analyste).

c. CREATION (date, all documents/Sectors) containing the date on which the document was created i.e. information at present in the fields :creation and :ver (for all documents created before the creation of the field :creation this field must be completed with the dates n found in the part MODI=n of the field :ver).

d. UPD.A (date, all documents/Sectors) containing the date on which the analytical part of the document was updated i.e. information at present in the fields :upd.a and :maj (for all documents updated before the creation of the field :upd.a this field must be completed with the dates m found in the part MAJO=m of the field :maj). For new documents not yet updated this date must be 0000/00/00.

e. UPD.T (date, all documents with textual fields) containing the date on which the textual part of the document was updated i.e. information at present in the fields :upd.t and :maj.txt (for all documents that had their textual part updated before the creation of the field :upd.t this field must be completed with the dates p found in the part = p of the field :maj.txt). For new documents not yet updated this date must be 0000/00/00.

f. MAJ.INFO (alphanumerical, text, all documents/Sectors) containing:

(i) the information at present in the second part of the field :maj.txt i.e. the number of the updating of the textual part of the document.

(ii) any other information that could be useful in the future for the members of the CELEX management team.

7. INFO will be a special field not associated with a lexicon for the display of information on the updatings of the base contained in the INFOCELEX document.

ANNEX 2D : INPUT CONTROL/PROCESSING SPECIFICATIONS FOR NEW CELEX'S FIELDS

abr. = abbreviated
tr. = translated
<...> = syntactical entity
; = descriptor separator
= paragraph separator
<...>ⁿ = n repetitions of the same entity possible
/ = or
* = and/or
{<...>/<...>} = different possible entities
[...]ⁿ = n repetitions of whole combination possible

FIELD NAME	INPUT IN ARCHIVE	CONTROLS	FORMAT IN ARCHIVE	TYPE IN BASE	FORMAT IN BASE	REMARKS
UMERO	.given by analyst	.position 1=Sector	<numero>	.alphanumeric	idem	.standardized identification number
ND (CCL)	.automatic generation for	.position 2,3=Year	SYTYNNNN(nm).m	.keywords		.all documents, all Sectors
ID (SCAD)	SC : documents in the context of text input	.position 4=TYPE .position 5-8= consistency with numero in TITRE	the part (nm).m is used in special cases			.max length 15 characters .no translation

IB

E (CCL)	.code given by analyst .partially automated generation on the basis of NUMERO	.consistency with NUMERO .controlled vocabulary	<abr. type> (with synonymy Long/Short)	.alphanumeric .keywords	<tr. type> (TABLES 030 & part of 301)	type of document .all documents, all Sectors .mean length 35 characters .max length 60 characters (for some languages) .introduction in ARCHIVE of abbreviated code synonymous with plain word .TYPE="suite" for overflow documents .translation of ARCHIVE code to all language versions .in the bases use of synonymy (Sector code letter) for searching
EUR (CCL)	.code (for Institutions & Organisations) or name given by analyst .automatic generation of group relation with Parliamentary Group if not specific .automatic generation of group relation for 1st or 2nd Legislature on the basis of the date	.consistency with NUMERO .controlled vocabulary	<abr. Institution ;>n {<person's name ;>n/ {<tr. Legislature ;> {<tr. Parliamentary/CES group ;>n<person's name ;>n/ {<tr. country ;>n (with synonymy and group relations)	.alphanumeric .keywords	<tr. Institution ;>n {<person's name ;>n/ {<tr. Legislature ;> {<tr. Parliamentary/CES group ;>n<person's name ;>n/ {<tr. country ;>n (TABLES 010,013,014,285, 290,130)	.author of document .all documents, all Sectors .mean length 17 characters .max length 60 characters (for some languages) .introduction in ARCHIVE of abbreviated code synonymous with plain-word code .translation of ARCHIVE code to all language versions .fusion of actual fields :auteur, :legislat and part of fields :groupe and :avoc.g .in the bases use of synonymy and group relations for searching
F.PUB 0 (CCL)	.given by analyst .automatic generation for some documents in the context of text input	.consistency with NUMERO .date format non redundant with page numbers	[<abr. publication type>; [<number>;<date>;<page>#]n .text	.alphanumeric	[<tr. publication type>; [N° <number>] ; du <date> ; p. <page>#]n (TABLES 210,306)	.publication reference of document .all Sectors .format of presentation in the bases automatically generated by Interface; can be changed by DBM .translation of ARCHIVE code to all language versions .14 publication types for existing documents .mean length 50 characters .max length 80 characters (?)

ESCRPT CT (CCL) CL (SCAD)	.code(s) or keyword(s) given by analyst	.consistency with existing thesaurus .controlled vocabulary	<abr. descriptor ;> ⁿ (with synonymy)	.alphanumeric .keywords	<tr. descriptor ;> ⁿ (TABLES 300,070,071)	.keyword classification of documents .all Sectors .fusion of actual fields :descript, :mat(ières) .use of standardized thesaurus .use of thesaurus updating procedures .in the bases use of synonymy and group relations for searching .translation of ARCHIVE code(s) to all language versions .mean length 200 characters .max length ?
QUER	.code (for Institutions) or name given by analyst	.consistency with NUMERO .controlled vocabulary	[<abr. Institution ;> ⁿ] [<abr. identification ;> ⁿ {<person's/><company's name ;>} ⁿ] [<abr. country> ⁿ]	.alphanumeric .keywords	[<tr. Institution ;> ⁿ] [<tr. identification ;> ⁿ {<persons'/company's name ;>} ⁿ] [<tr. country> ⁿ]	.name(s) or type(s) of applicant(s), defendant(s), observer(s) or intervenant(s) in case(s) before the Court .Sector 6 .translation of ARCHIVE code to all language versions .mean length 17 characters .max length 60 characters (for some languages)
FEND SERV ITERV			[<abr. Institution ;> ⁿ] [<abr. identification ;> ² {<person's/><company's name ;>} ⁿ] [<abr. country ;> ⁿ]		(TABLE 110 120 160)	
ITIONA	.code given by analyst or automatic generation on the basis of name of MEP (use of thesaurus)	.consistency with thesaurus .controlled vocabulary	<abr. country ;> ⁿ (with synonymy)	.alphanumeric .keywords	<tr. country ;> ⁿ (TABLES 140,290)	.nationality of applicant, defendant, national court or Parliamentarian .Sectors 6,9 .translation of ARCHIVE code to all language versions .mean length 20 characters .max length 80 characters (for some languages) .in the bases use of synonymy (with ISO country codes) for searching

B. NAI	.free text given by analyst .consistency with NUMERO	free text	.alphanumeric .text	free text	.references to proceedings before the national courts in the context of a preliminary ruling .Sector 6 .no translation .mean length 350 characters .max length ?
I. G I. RAP	.name given by analyst .consistency with NUMERO .controlled vocabulary	<person's name>	.alphanumeric .keywords	<person's name> (TABLE 130)	.name(s) of Advocate general Judge-Rapporteur of Case .Sector 6 .translation of ARCHIVE code to all language versions .mean length 17 characters .max length 60 characters
JEDUR	.code given by analyst .consistency with NUMERO .controlled vocabulary	<abr. procedure> (with synonymy)	.alphanumeric .keywords	<tr. procedure> (TABLE 100)	.nature of proceedings .Sector 6 .translation of ARCHIVE code to all language versions .mean length 45 characters .max length ? .in the bases use of synonymy for searching
ERT	.code given by analyst .consistency with DESCRIPT .controlled vocabulary	<numerical descriptor> ⁿ (with synonymy) <i>group</i>	.alphanumeric .keywords	[<numerical descriptor ;> <tr. descriptor ;>] ⁿ (TABLES 090, 070, 071)	.classification according to codes of Directory of Community legislation in force .Sectors 1-7 .translation of ARCHIVE code to all language versions .in the bases use of synonymy and group relations for searching .use of thesaurus updating procedures .mean length 70 characters .max length ? .fusion of codes in actual field :repet and of their translation in field :mat

ANGUE .A (CCL)	.code given by analyst	.controlled vocabulary	<abr. language ;> ⁿ (with synonymy)	.alphanumeric .keywords	<tr. language ;> ⁿ (TABLE 060)	.language of proceedings .Sectors 1-4,6-8,10 .translation of ARCHIVE code to all language versions .in the bases use of synonymy for searching .mean length 30 characters .max length 80 characters
:STINA	.code or name given by analyst	.controlled vocabulary	[<abr. Institution ;> ⁿ] [<abr. identification ;> ² <person's/company's name ;> ⁿ [<abr. country> ⁿ] (with synonymy)	.alphanumeric .keywords	[<tr. Institution ;> ⁿ] [<tr. identification ;> ² <person's/company's name ;> ⁿ [<tr. country> ⁿ] (TABLE 050)	.names of person(s)/Institution(s) to which the act is addressed .destinators of food aid .S 2-5 .translation of part of ARCHIVE codes in all language versions .in the bases use of synonymy for searching .use of thesaurus updating procedures .mean length 35 characters .max length ?
:POS	.code given by analyst	.controlled vocabulary .consistency with NUMERO	<abr. Institution> (with synonymy)	.alphanumeric .keywords	<tr. Institution> (TABLE 040)	.depository body for agreements .Sector 2 .translation of ARCHIVE code to all language versions .in the bases use of synonymy for searching .mean length ? .max length 80 characters
SP. XX XX=FR,UK,NL, DE,DA,IT,ES, PO,EL,BE,LU, IR)	.input by ASHODEE .comments given by analyst ?	.free text .consistency with NUMERO .controlled format .data consistency controls	free text	.alphanumeric .text	free text	.national provisions implementing directives .Sector 3L .mean length ? .max length ? .no translation .standardised phrase for directives with no need for national measures of implementation (now in field :info.div) .max length ?

:SP :SO	.code given by analyst	.controlled vocabulary	<service> ⁿ	.alphanumeric .keywords	<service> ⁿ	.department responsible associated .Sectors 3-4 .no translations .mean length Resp = 10, Asso = 30 .max length Resp = 9, Asso = 45
:SE	.numero and other information given by analyst	.consistency with NUMERO .proper format .controlled vocabulary	[<numero> [AR-<n> ⁿ [AL-<m> ⁿ]]] ⁿ (with synonymy)	.alphanumeric .text	[<numero> [Article-<1> [#] Article-<n> ⁿ [Alinea-<m> [#]]]] ⁿ (TABLE 370)	.legal base of document .Sectors 2-7 .translation of standard part of ARCHIVE code in all language versions .use of synonymy in the bases for searching treaties .use of thesaurus updating procedures for treaties (actual table 020) .mean length ? .max length ?
:IT	.numero and other information given by analyst	.consistency with NUMERO .proper format .controlled vocabulary	[<numero> [AR-<n> ⁿ [AL-<m> ⁿ]] [N-<i> ⁿ][P-<o> ⁿ]] ⁿ	.alphanumeric .text	[<numero>[Article-<1> [#] Article-<n> [Alinea-<m> [#]] [N-<i> ⁿ][P-<o> [#]]]] ⁿ (TABLE 371)	.acts cited in the document .all Sectors .translation of standard part of ARCHIVE code in all language versions .mean length ? .max length ?

MOD .numero and other infor-
 given by analyst
 .consistency with
 NUMERO <numero>{<abr. ac.mod.type>- .alphanumeric
 <abr. part of doc. affected> .text with
 .proper format [<number>] [FR-<date> TL- positional
 .controlled <date>];]n]n indexation
 vocabulary
 .acts amended by document
 .all Sectors
 .for amendments of the same type to different
 articles of the same document (i.e. same
 id.number) the id.number is given only once.
 .data are ordered according to ascending id.number
 .translation of standard parts of ARCHIVE code in
 all language versions
 .fusion of actual fields :act.mod, :act.vis,
 act.jur, :ant.jur
 .mean length ?
 .max length ?
 (TABLES 275,375,395,377)

ILI .automatically generated
 on the basis of MOD
 .automatically erased if an
 entry in MOD is erased
 .proper format <abr. part of doc. affected> .alphanumeric
 .controlled [[<number>]{<tr.pas.mod.type> .text
 vocabulary <numero> [FR-<date> TL-<date>];]n]n
 <numero> [FR-<date> TL-<date>]#]n]n .for more than one amendment of the same article
 by the same document, the article and its number
 are given only once
 .data are ordered according to ascending article
 number
 .translation of standard part of ARCHIVE code in
 all language versions
 .fusion of actual fields :mod.ult, :vis.ult,
 :arrêts
 .mean length ?
 .max length ?
 (TABLES 270,375,396)

A. PREP	.numero or code given by analyst	.consistency with NUMERO controlled vocabulary .automatic generation of TYPE on the basis of NUMERO .proper format	[[{<abr. tra.prep>}{<numero>}{<reference>}]] ⁿ (with synonymy)	.alphanumeric .text (TABLES 390,680)	[[{<tr. tra.prep>}{<numero>}{<reference>}]] ⁿ	.information on preparatory work .all Sectors .translation of standard part of ARCHIVE code in all language versions .use of synonymy in the bases for searching .mean length 60 characters .max length ?
TES	.free text given by analyst	.free text	free text	.alphanumeric .text	free text	.references of articles in relation with a case before the Court .Sector 6 .no translation .mean length ? .max length ?
<u>LAT</u> idden from users)						
S.1	-	-	-	.alphanumeric .keywords	<numero> ⁿ	.id.numbers of documents forming the legal base of the act .Sectors 2-7 .generated automatically from field BASE of the ARCHIVE .hidden from users; to be used in the context of a macrocommand for Selection Deduite questions
T.1	-	-	-	.alphanumeric .keywords	<numero> ⁿ	.id.numbers of documents cited in act .all Sectors .generated automatically from field CIT of the ARCHIVE .hidden from users to be used in the context of a macrocommand for Selection Deduite

IE

C .date given by analyst <date> .date
 .format control <date> .date
 .consistency

1 .date given by analyst <date> .date
 .format control <date> .date
 .consistency

ET .date and code given by analyst <date> <abr. date code ;>]: .date with characteristics
 .controlled <date> <tr. date code ;>]:
 vocabulary (TABLES 335,336)

VAL .date given by analyst <date> .date
 .format control <date> .date
 .consistency

NSPOS .date given by analyst <date> .date
 .format control <date> .date
 .consistency

.date of document
 .all Sectors
 .no translation
 .mean length 10 characters
 .max length 10 characters

.date the application was lodged
 .Sector 6
 .no translation
 .mean length 10 characters
 .max length 10 characters

.date the act enters into force
 .Sector 1-5
 .translation of ARCHIVE codes in all language versions
 .mean length 42 characters
 .max length 90 characters
 .fusion of actual fields :effect and :notif

.date of end of validity
 .Sectors 1-5
 .no translation
 .dates of reconduction of agreements (Sector 2) actually in field :info.div
 .mean length 10 characters
 .max length 10 characters

.date by which the Member States can comply with the directive
 .Sector 3
 .no translation
 .mean/max length 10 characters

310N	.date and code given by analyst	.format control	[<date> <abr. date code ;>]?	[<date> <tr. date code ;>]?	.date or corresponding session and voting of the E.P.
		.consistency			.Sectors 5,9
		.controlled vocabulary			.translation of ARCHIVE codes in all language versions
					.mean length 20 characters
					.max length 23 characters
RE	.French title automatically fed in by TEXTERFACE	.existence	free text	free text	.title of document; obligatory in bases
	.possibility to introduce title directly by analyst	.for some types of doc's consistency with NUMERO		.positional	.all documents, all Sectors
					.for overflow documents = special title
					.for documents without title (especially in non-French version) = special title
					.special arrangements for introducing titles in non-French bases (Texterface, direct introduction with special characters)
					.mean length 210 characters ?
					.max length ?
A	INDEX	(not treated in ARCHIVE)		free text	.text of documents
SID	SOMMAIRE				.automatic field separation by TEXTERFACE
PUS	INTITULE				.automatic generation of overflow fields/documents
TNOTE	MOTIFS				
EXE	DEPENS				
IIF	VISA				
	DISPOSI				
	CONCLUS				
	RESUME				

```

FO.DIV .free text given by analyst .free text .free text .alphnumeric free text .misc. information
.Sectors 2-7
.fusion of fields :info.div and part of fields
:ref.int, :comment, :index.cr
.no translation
.actual TABLES 400,412,301
.mean length 60 characters

) .code given by analyst .controlled <rep>? <rep>? .alphnumeric <rep>? .information for producing the Directory
.vocabulary (TABLE 091)
.no translation
.Sectors 3,4
.mean length 3 characters
.max length 6 characters

LYSIE .automatically fed by <analyst ;>n <analyst ;>n .alphnumeric <analyst ;>n .name of analyst
machine = USERID (?) .controlled <analyst ;>n .keywords .no translation .all Sectors .mean length ? .max length ?
.no translation

ATION .automatically fed by .format control <date> <date> .date .date of creation of document
machine .format control <date> <date> .all doc's/all Sectors .no translation .mean/max length 10 characters

.A .automatically fed by .format control <date> <date> .date .date of updating the analytical part of document
machine .format control <date> <date> .all doc's/Sectors .no translation .mean/max length 10 characters
.for new documents 0000/00/00

```

JPD.I	.automatically fed by machine	.format control	<date>	<date>	.date of updating the textual part of document .doc's with textual part .no translation .mean/max length 10 characters .for new documents 000/00/00
IAJ.INFO	.automatically fed by TEXTERFACE .free comment by analyst	-	free text	free text	.useful information concerning the document .mean length ? .max length ? .no translation
INFO	.automatically fed by INFOCELEX stream in the base	-	-	free text	.information on the updating of the base .mean length ? .max length ?

Annex 2E Types of "bordereaux d'analyse" screens originally needed for the Archive input.

- Sector 1 - Treaties
- Sector 2 - Agreements
- Sector 3 - Regulations
 - Directives
 - Decisions
 - Other acts (including Sector 4 documents)
- Sector 5 - Commission proposals
 - EP documents
 - Documents of other Institutions (ESC, Court of Auditors)
- Sector 6 - Judgements
 - Opinions of Advocates-General
 - New cases introduced and rejected
- Sector 8 - ?
- Sector 9 - Parliamentary questions
- Sector 10- ?



Bruxelles, le 27 janvier 1988

VB/ch

-
IX.F.4 - S.I.I.

-
Bases de données

-
(89) D/26

Note à l'attention de Monsieur DE BACKER

Objet : Relevées des modifications apportées aux documents "Préanalyse" et "Aspects utilisateurs"

Référence : Réunion de M. DE BACKER, Mme SCHOLZ et M. BROCHART du 2 décembre 1988

I. DOCUMENT "PREANALYSE"

Relevé des modifications incorporées dans la version de novembre 1988 du "Projet de préanalyse pour le développement et la modernisation du système CELEX" :

- Point III.C.2 (page 29) : Relation avec IDS
Distinction entre les relations conceptuelles et les relations techniques (outils);
- Point III.C.3.2 (page 29) : Solutions alternatives
Risques présentés par l'installation de la base de production des textes sur machine locale;
- Point III.C.3.3 (page 30) :
Logiciels pris en compte pour l'étude faisabilité
- Point V.A (page 35) : Stratégie de développement
Séparation du projet global en deux sous-projets distincts :
 - production des textes;
 - production des données analytiques, diffusion et "module pilote"
- Point V.B (pages 36 à 46) : Scénario de mise en place
Description détaillée des trois différentes solutions envisageables pour la mise en place du nouveau système;
- Point V.C (page 48) : Ressources
Réévaluation des ressources nécessaires, compte tenu des modifications mentionnées ci-dessus et de la réalisation d'une maquette au cours de la phase "étude de faisabilité";



- Point VI (pages 49-50) : Etude de faisabilité
Redéfinition des éléments de l'étude de faisabilité, compte tenu de la séparation en deux sous-projets et de la réalisation de la maquette.

II. DOCUMENT "UTILISATEURS"

1. Adjonction au document "Aspects utilisateurs" d'une annexe décrivant les nouveaux champs CELEX et incorporation dans le document de la référence à cette annexe.

- a) Point IV.A (page 13) : Structure
Ajout de la phrase suivante en fin de page : un tableau des nouveaux champs et lexiques figure en annexe 1.

- b) Cette nouvelle annexe est jointe au présent document.

2. Incorporation d'une mention relative à EUROVOC

Point IV.A (page 12) : Structure
Ajout de la footnote (1) suivante :

(1) Il est envisagé de généraliser l'implantation du thésaurus EUROVOC 2 dans CELEX si l'expérience d'introduction progressive dans le champ DESCRIPT actuellement en cours au Parlement européen est favorablement accueillie par les utilisateurs.

3. Adjonction d'une annexe décrivant les grands points du CCL

- a) Point IV.B.2 (page 15) : Le CCL
Ajout de la phrase suivante à la fin du point 2 :
Une brève description du CCL figure en annexe 2

- b) Cette annexe est jointe au présent document.

4. Précision sur le nombre et l'architecture des menus

Point IV.B.3 (page 17) : L'interface conviviale

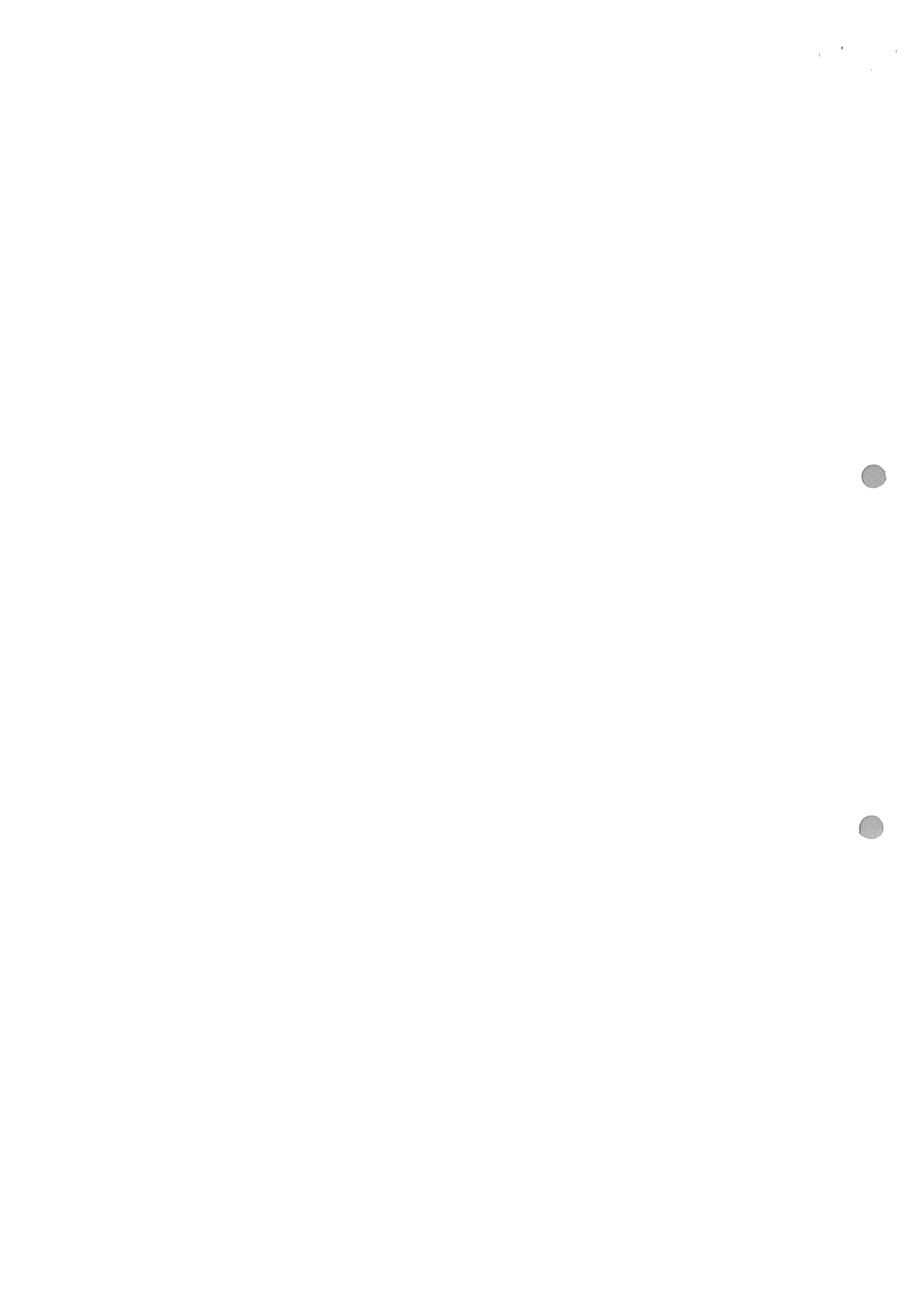
Ajout de la phrase suivante dans le 3^{me} paragraphe avant "Par ailleurs ..." :

"A cette fin les menus proposés seront peu nombreux (4 à 6) tout en présentant chacun un grand choix de stratégies de recherche.

5. Explication de l'abréviation UAS

Point IV.B.3 : L'interface conviviale, à la page 18 dans la figure 1, la footnote (1) est ajoutée à l'abréviation UAS :

(1) USER AGENT SERVICE dont l'objectif général est l'amélioration de l'accès aux services informatiques pour la création d'outils tels que user agent (guide des services), interfaces conviviaux de recherche (UFI), système de type vidéotex (infotex), outils de transfert et de reformatage, etc.



6. Allusion à l'objectif de réduire au maximum les besoins de matériel pédagogique

Point VI.C (page 20) : Documentation et matériel pédagogique

Introduction après le titre du point C du paragraphe suivant :

L'idéal serait évidemment de pouvoir rendre le système suffisamment simple pour que l'utilisateur n'ait besoin d'aucun manuel. Comme cependant exposé plus haut, la réalité complexe du droit communautaire exclut une approche aussi radicale. Il paraît par conséquent utile de parvenir à déterminer avec précision les besoins réels des utilisateurs en la matière, en testant dans un premier temps l'utilisation du prototype sans aucun support pédagogique, et de ne fournir que dans un deuxième temps la documentation adaptée aux besoins exprimés par les utilisateurs au cours de la première phase. Dans l'attente des résultats concrets de ce test, une approche pragmatique et traditionnaliste s'impose.

7. Introduction d'un résumé du projet plan et d'une évaluation chiffrée des bénéfices dans le chapitre IV

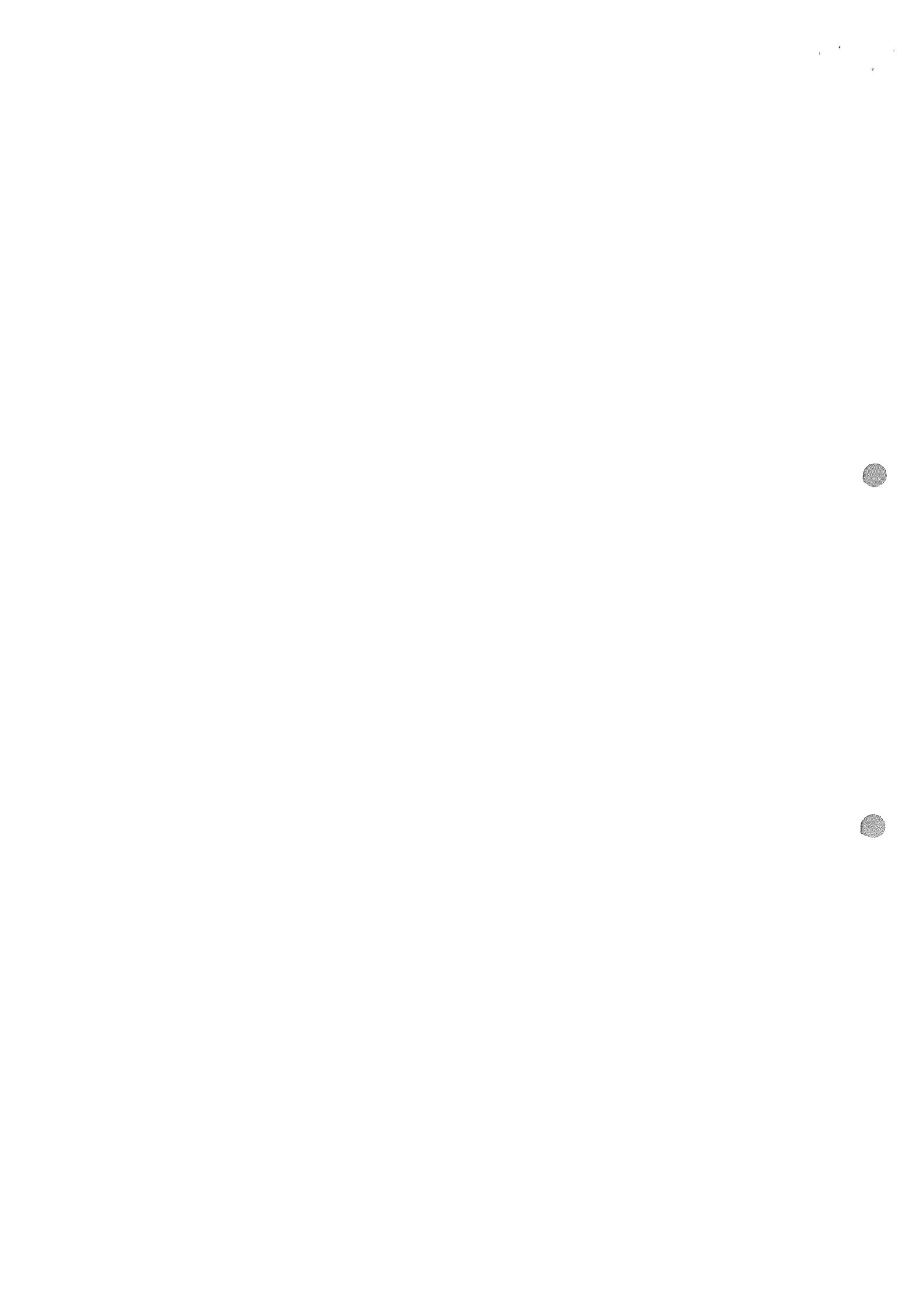
Le chapitre IV (page 23), "Les bénéfices pour l'utilisateur final" est modifié comme suit :

- a) introduction d'un nouveau point A) "Mise en place de la nouvelle version de CELEX", qui présente sous forme de synthèse un résumé du project plan vu par l'utilisateur;
- b) incorporation des trois paragraphes restants du chapitre dans un nouveau point B) :
"Evaluation de la réduction des coûts d'utilisation", qui chiffre les bénéfices attendus de la modernisation tout au long du processus d'utilisation."

Le nouveau chapitre VI est joint au présent document

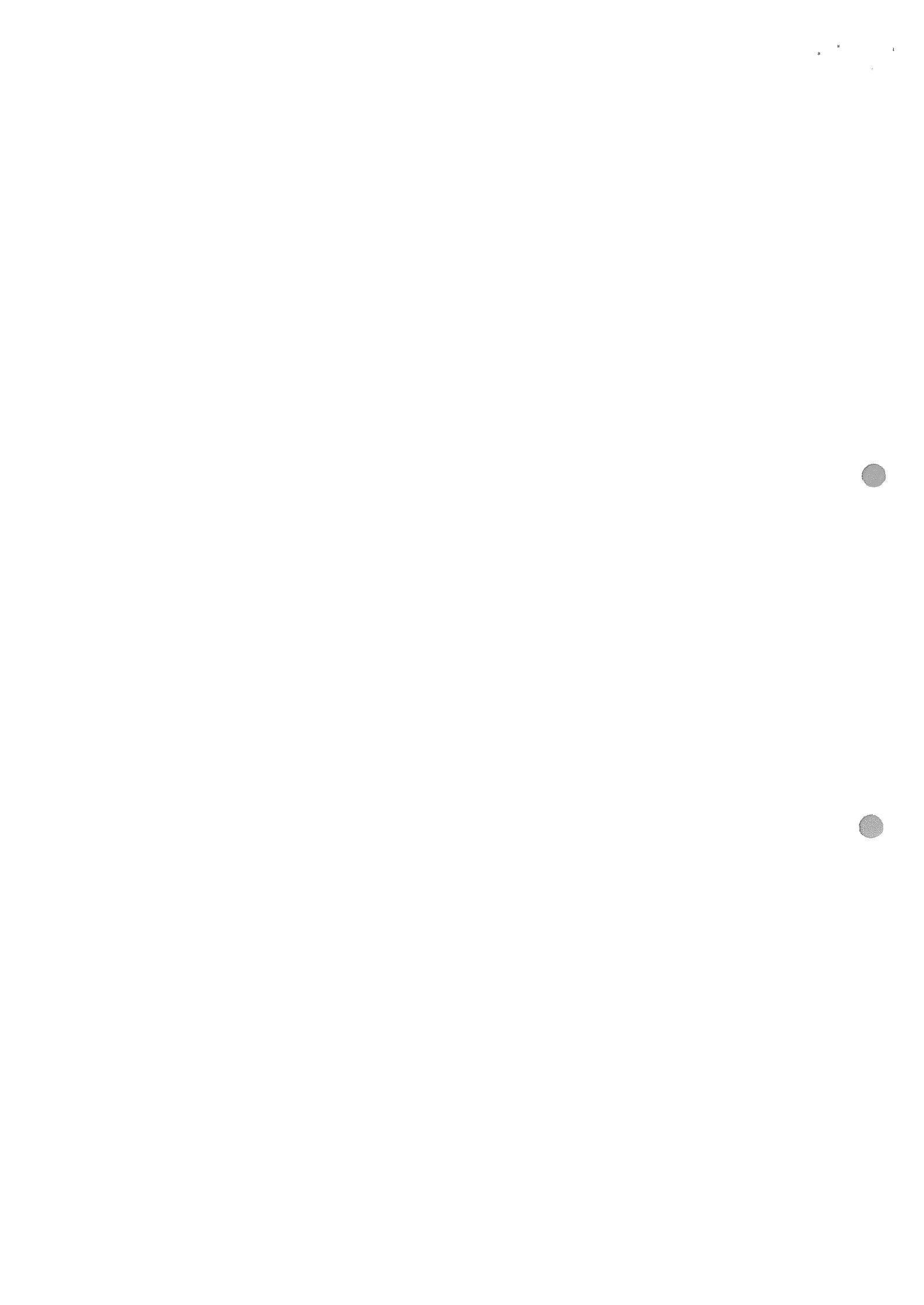

P. BROCHART


V. BENSCH

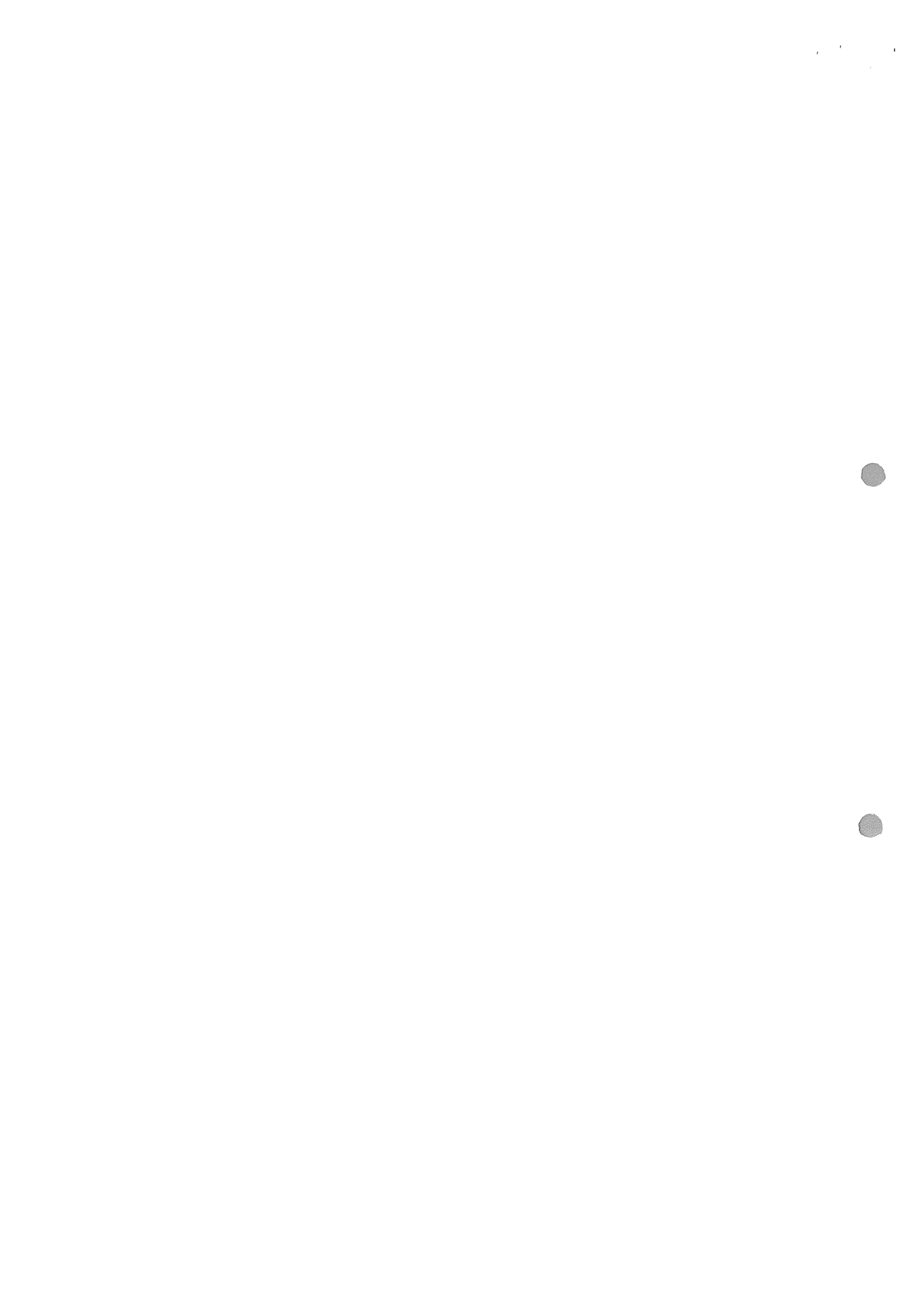


NOUVEAUX CHAMPS ET LEXIQUES CELEX

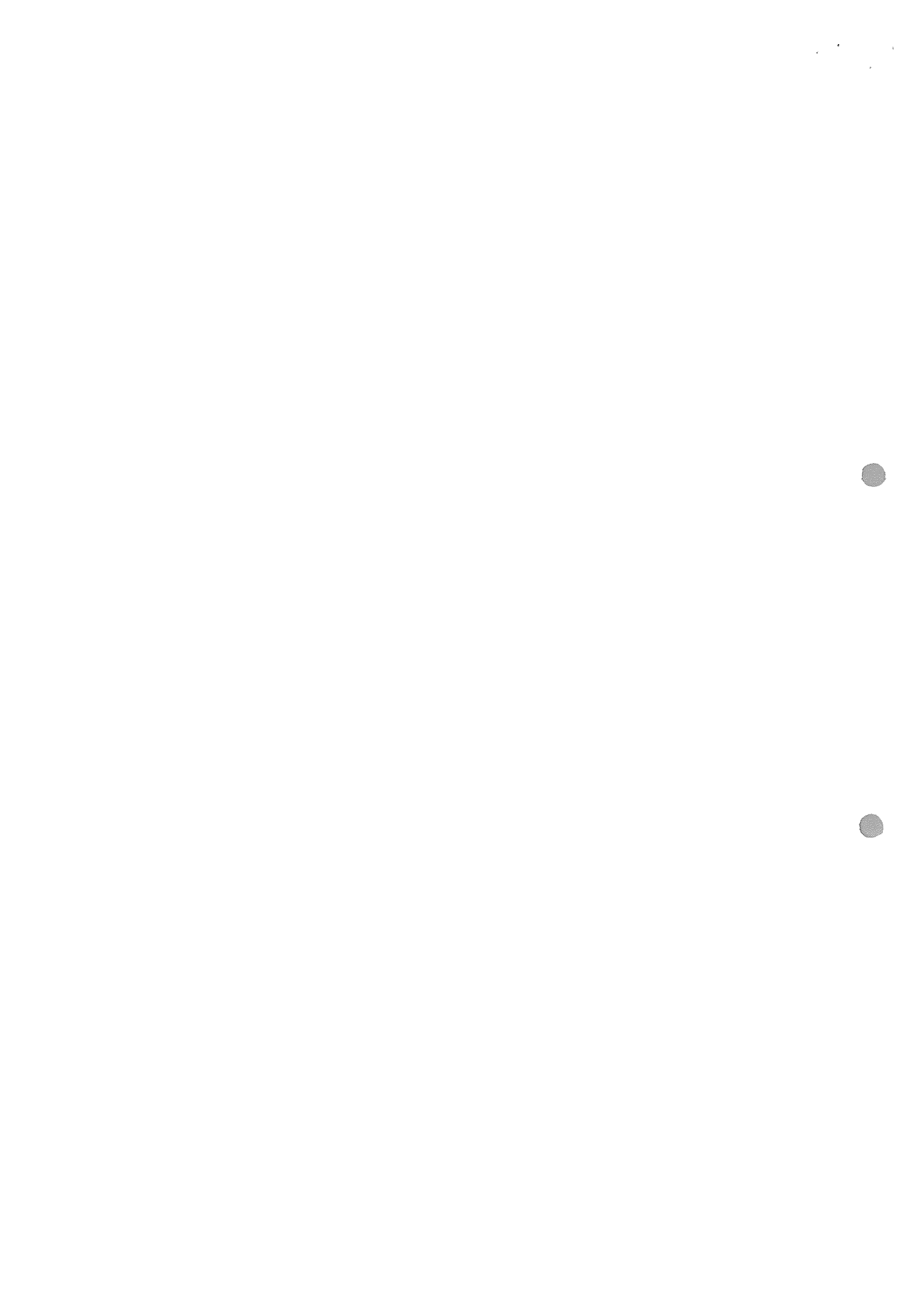
NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB	NUMERO	ND	Numéro d'identification standardisé du document (SAATNNNN(nn))	TOUS SECTEURS ¹ Alphanumérique, vocabulaire contrôlé
	TYPE	DT	Nature juridique du document symbolisée par une lettre (R = Règlement, D = Décision)	TOUS SECTEURS Alphanumérique, vocabulaire contrôlé Synonymie entre la lettre du numéro et le libellé long de la forme juridique de l'acte
	AUTEUR	AU	Nom de l'institution, organe, pays dont émane l'acte; nome de la personne physique (parlementaire, Avocat général, etc.) auteur du document	TOUS SECTEURS Alphanumérique, vocabulaire contrôlé Synonymie et relation de groupe
	REF.PUB	SO	Références de publication au J.O. ou au RECUEIL (AAAA/MM/JJ)	TOUS SECTEURS Alphanumérique, texte



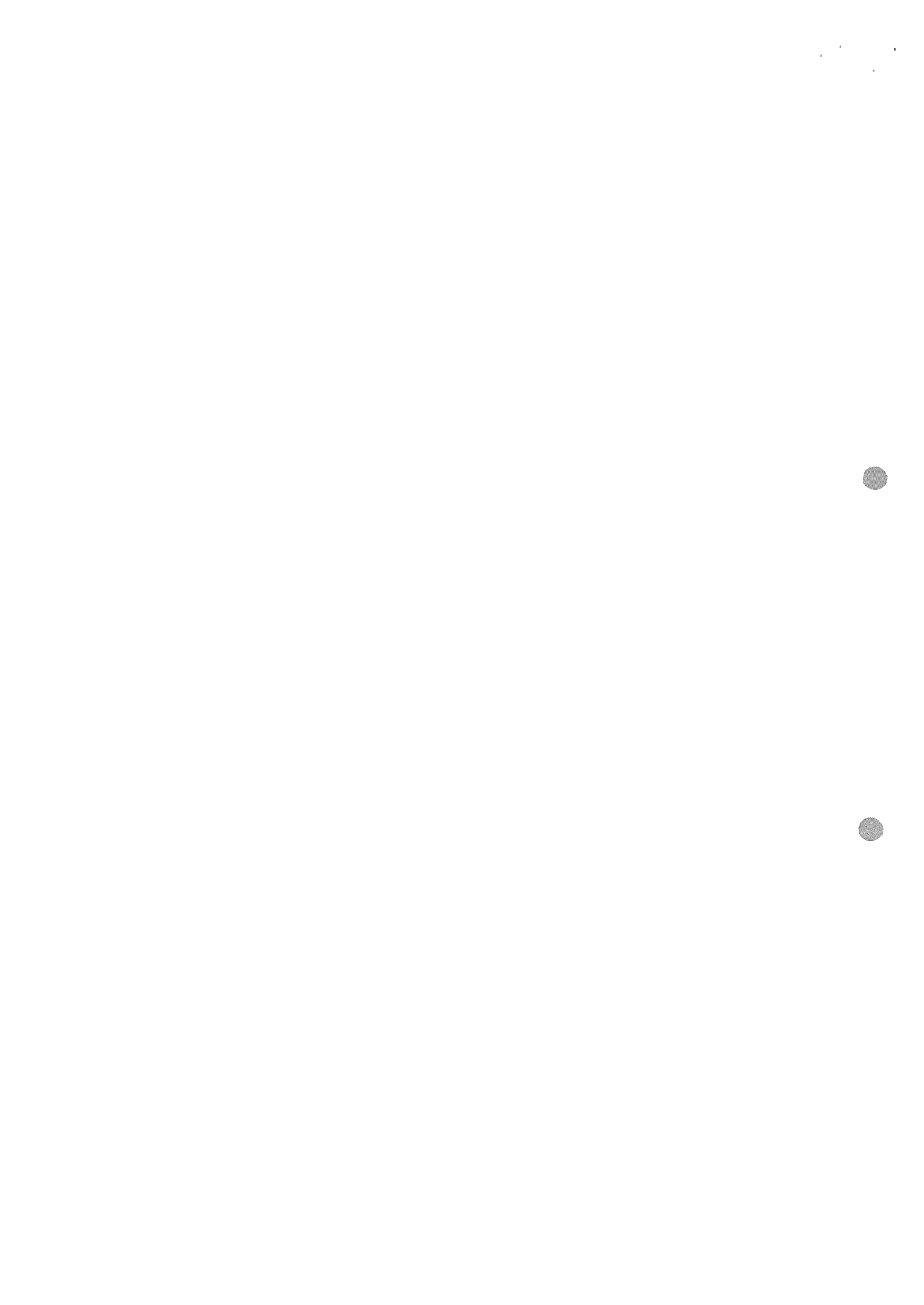
NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	DESCRIPT	CT	Descripteurs du contenu du document	TOUS SECTEURS Alphanumérique, vocabulaire contrôlé Synonymie et relation de groupe Thésaurus standardisé (peut-être EUROVOC ?)
	REQUER	-	Nom de la partie requérante dans une affaire devant la C.J. (institution, organe, pays, personne physique ou morale)	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	DEFEND	-	Nom de la partie défenderesse	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	INTERV	-	Nom de la partie intervenante	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	OBSERV	-	Institution, organe ou état membre ayant formulé des observations dans une affaire devant la C.J.	SECTEUR 6 Alphanumérique, vocabulaire contrôlé



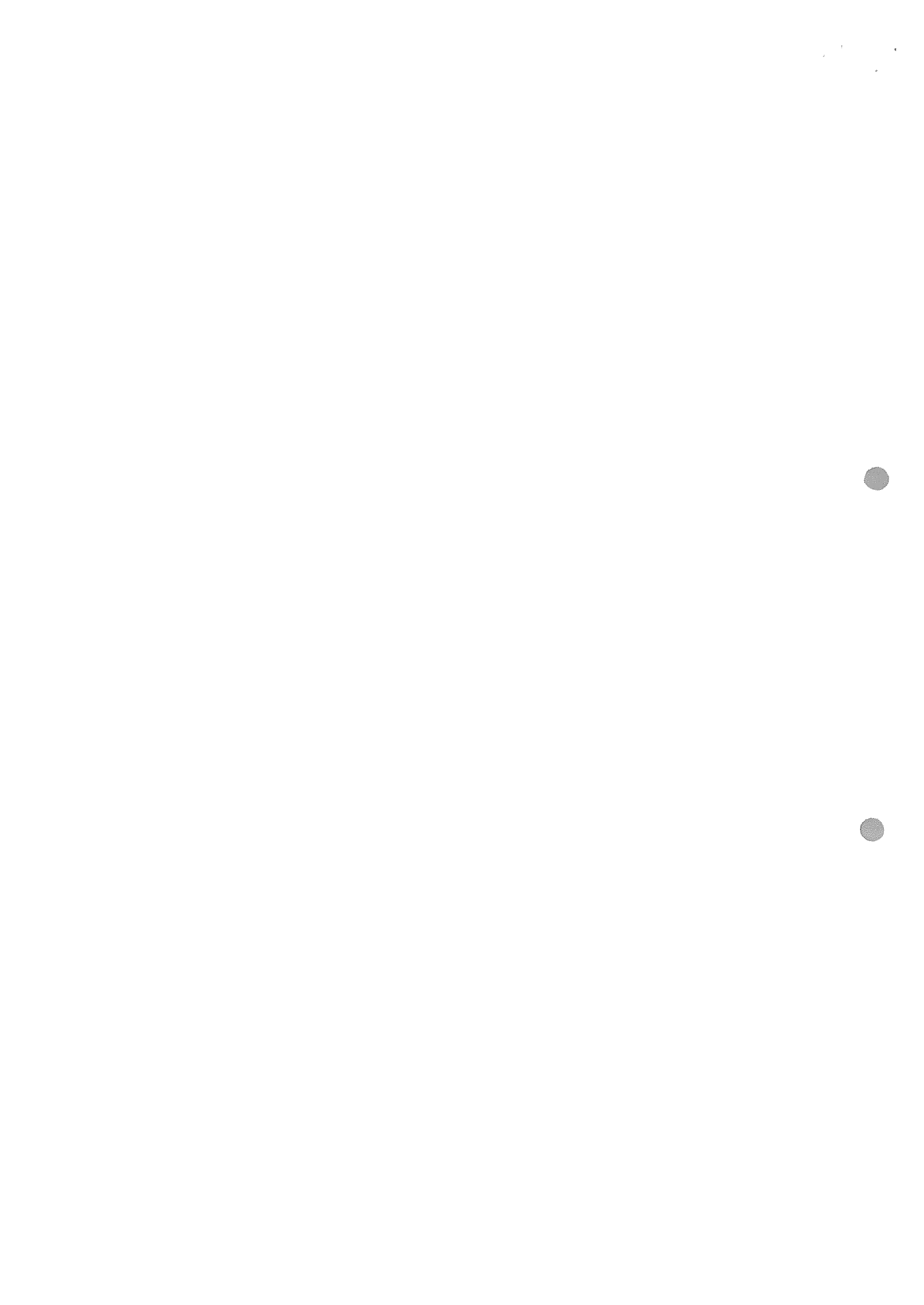
NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	NATIONA	-	Nationalité du requérant ou défendant particuliers ou du membre du PE Pays d'appartenance du tribunal dont émane la demande de décision pré-judicielle	SECTEURS 6 et 9 Alphanumérique, vocabulaire contrôlé Synonymie
	TRIB.NAT	-	Références des jugements des tribunaux nationaux ayant donné lieu à un renvoi pré-judicielle	SECTEUR 6 Alphanumérique, texte
	AVOC.G	-	Nom de l'avocat général	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	JUGE.RAP	-	Nom du juge rapporteur	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	PROCEDUR	-	Nature du recours, par exemple : (recours de fonctionnaire; recours en annulation, en carence, en constatation de manquement, demande en intervention, etc.)	SECTEUR 6 Alphanumérique, vocabulaire contrôlé Synonymie



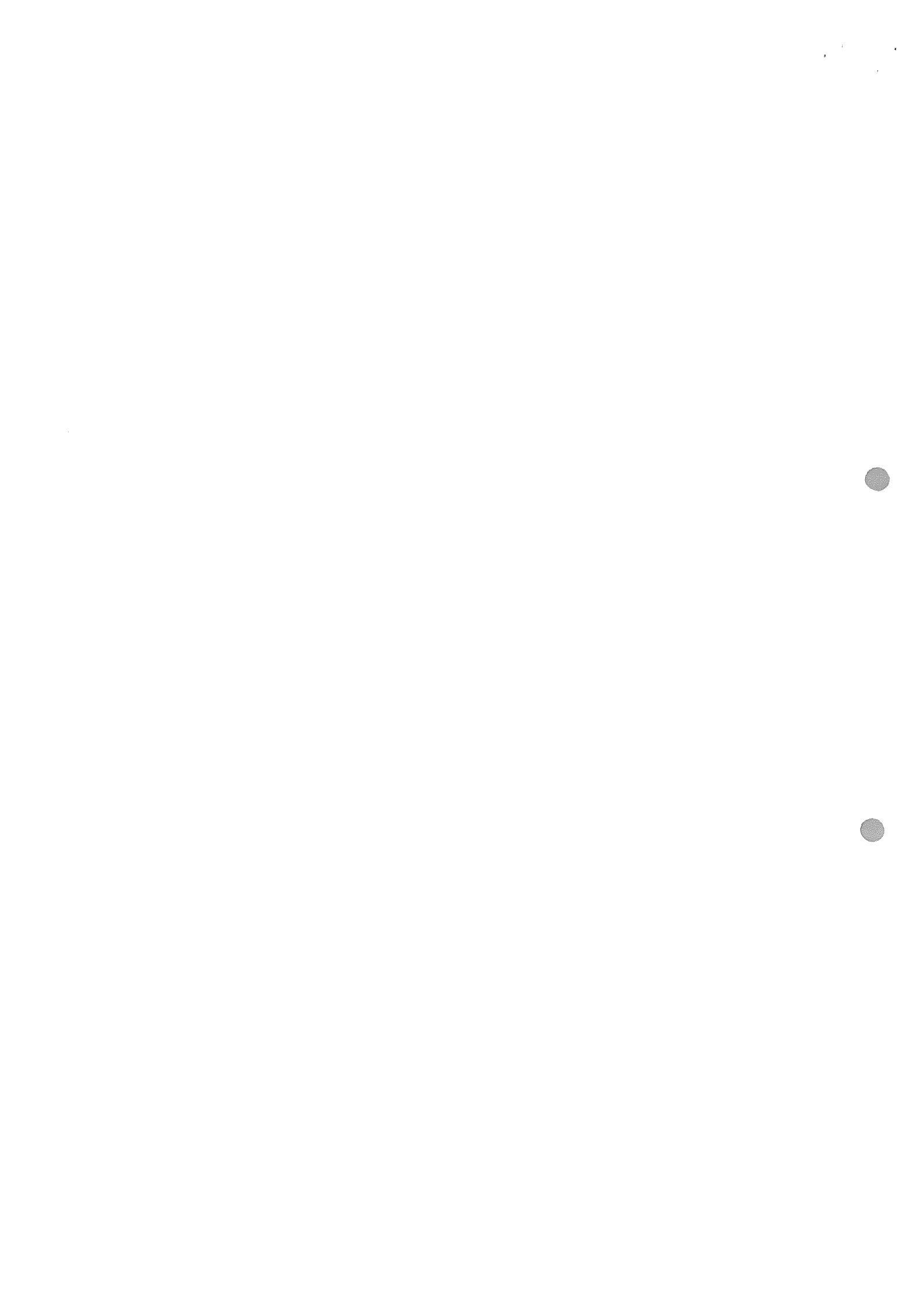
NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	REPERT	-	Code numérique de classification dans le "Répertoire de la législation communautaire en vigueur"	SECTEUR 1 à 7 Alphanumérique, vocabulaire contrôlé Synonymie entre le code numérique et les descripteurs correspondant à la classification par matières de l'activité législative Relation de groupe
	LANGUE	LA	Langue(s) faisant foi; langue de procédure	SECTEURS 1 à 4, 6 à 8, 10 Alphanumérique, vocabulaire contrôlé Synonymie
	DESTINA	-	Etats, institutions ou particuliers auxquels l'acte est adressé	SECTEURS 2 à 5 Alphanumérique, vocabulaire contrôlé Synonymie
	DEPOS	-	Organe dépositaire des accords	SECTEUR 2 Alphanumérique, vocabulaire contrôlé Synonymie
	DISP XX	-	12 champs contenant les dispositions nationales d'exécution des directives communautaires (XX = BE, DA, DE, EL, ES, FR, IR, IT, LU, NL, PO, UK) avec références de publication dans les JO nationaux	SECTEUR 3L Alphanumérique, texte



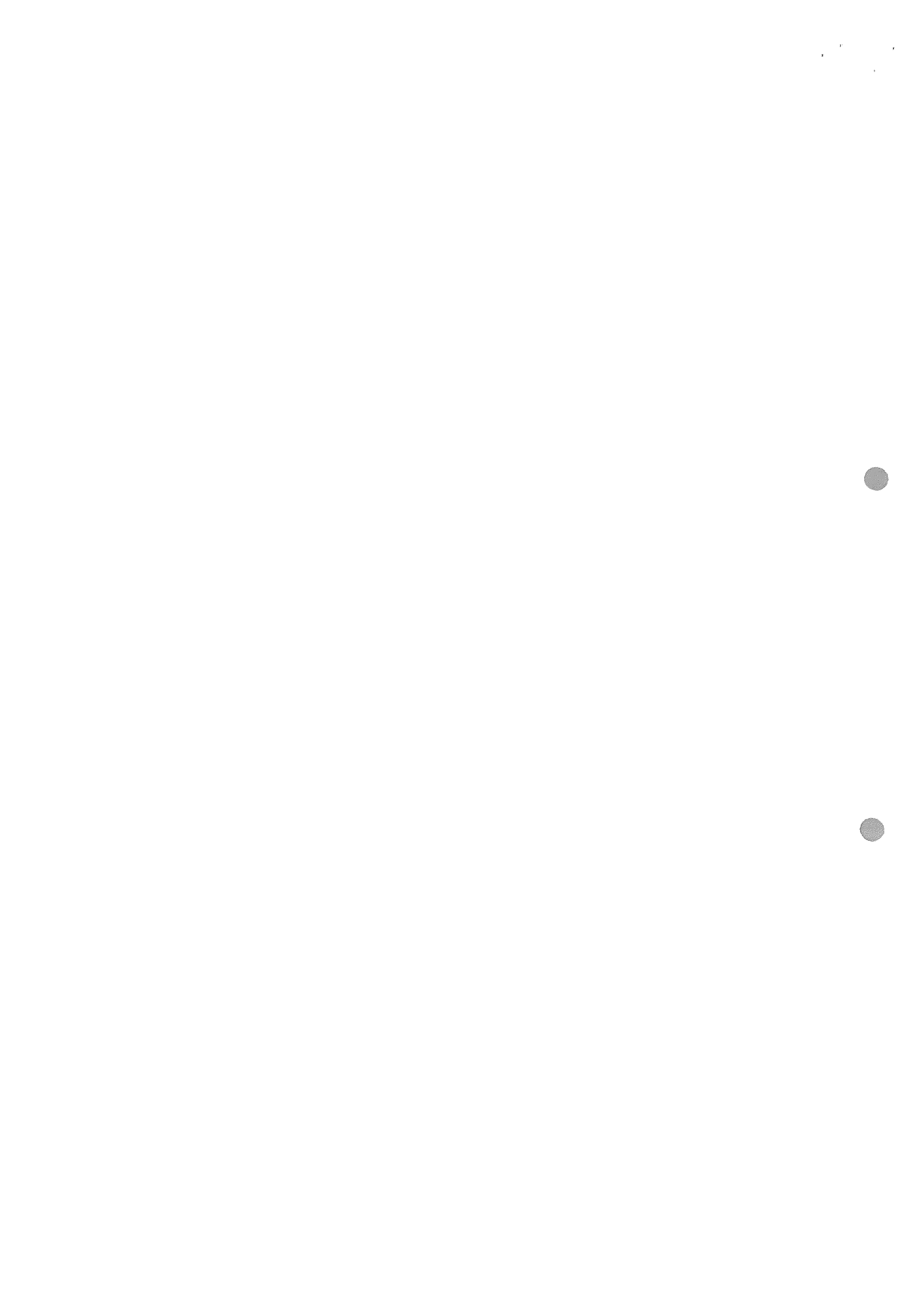
NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	INFO.DIV	-	Informations diverses, références internes des documents, commentaires utiles	SECTEURS 2 à 7 Alphanumérique, vocabulaire contrôlé
REL	BASE	-	Numéro d'identification des documents constituant la base juridique de l'acte examiné	SECTEUR 2 à 7 Alphanumérique, texte Synonymie
	CIT	-	Numéro d'identification des documents cités dans l'acte concerné	SECTEURS 2 à 7 Alphanumérique, texte
	MOD	-	Actes cités, visés, modifiés ou affectés par l'acte concerné	SECTEUR 1 à 7 et 9 Alphanumérique, texte avec indexation positionnelle
	ULT	-	Actes qui visent, modifient ou affectent l'acte concerné	SECTEUR 1 à 7 et 9 Alphanumérique, texte
	TRA.PREP	-	Références des actes préparatoires (propositions, avis) qui sont à l'origine de l'acte concerné	SECTEUR 1 à 7 Alphanumérique, texte
	NOTES	-	Références aux articles parus dans des revues juridiques relativement à l'arrêt concerné	SECTEUR 6 Alphanumérique, texte



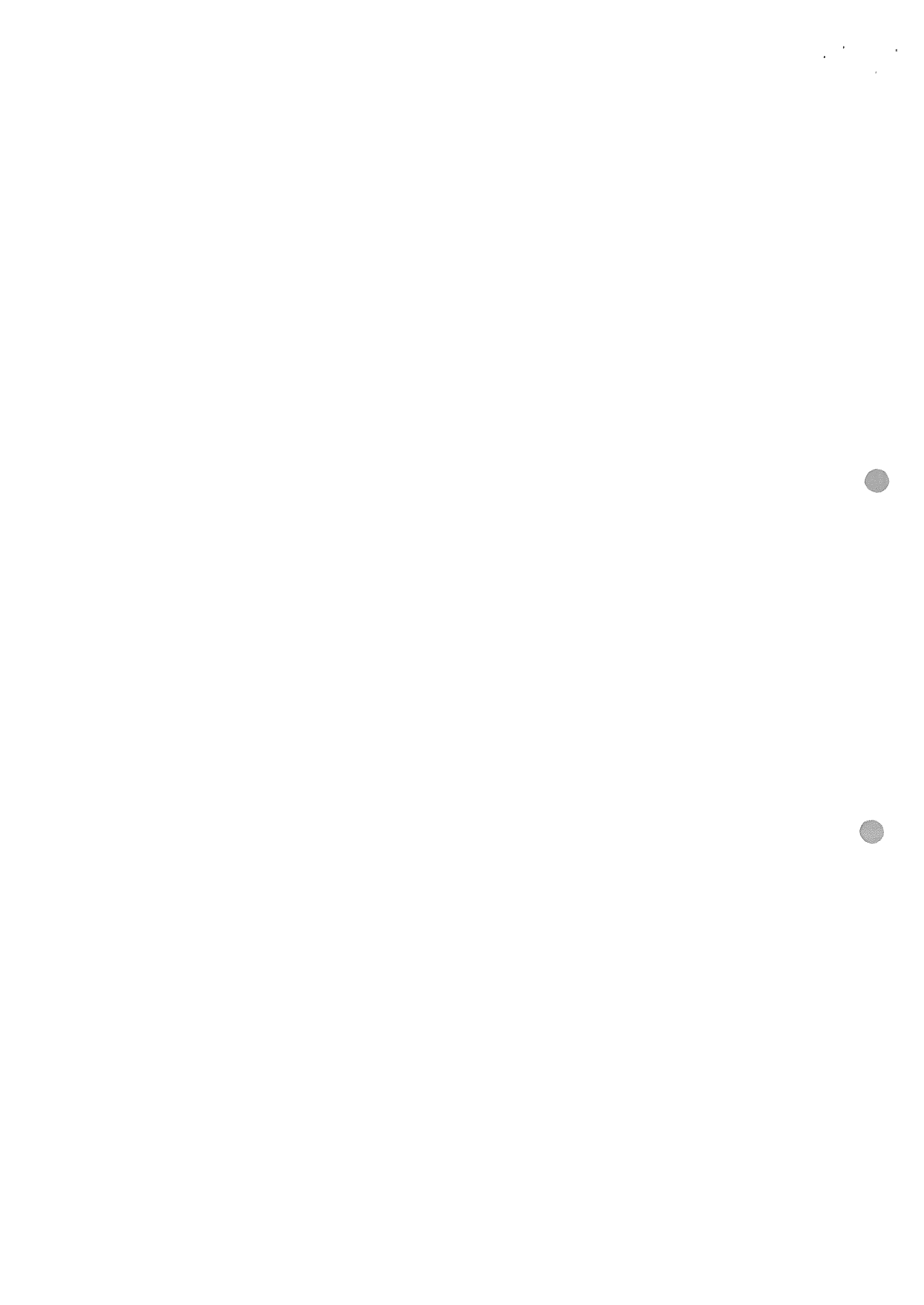
NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
DATE	DOC	-	Date d'adoption de l'acte, date du prononcé de la décision de la Cour (YYYY/MM/DD)	TOUS SECTEURS Date
	DEM	-	Date de dépôt d'une affaire au Greffe de la Cour (YYYY/MM/DD)	SECTEURS 6 et 8 Date
	EFFET	-	Date d'entrée en vigueur (YYYY/MM/DD ENTREE EN VIGUEUR) Date de mise en application (YYYY/MM/DD IMPLEMENTATION) Date de notification (YYYY/MM/DD NOTIFICATION)	SECTEURS 1 à 5 Date, avec caractéristique
	FIN.VAL	-	Date de fin de validité	SECTEUR 1 à 5 Date
	TRANSPOS	-	Date à laquelle les Etats membres doivent se conformer à une directive	SECTEUR 3 Date
	SESSION	-	Date du débat d'une résolution ou d'un avis du PE (YYYY/MM/DD DEBAT) Date de vote d'une résolution ou d'un avis (YYYY/MM/DD VOTE)	SECTEURS 5 et 9 Date, avec caractéristique



NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
TXT	TITRE	TI	Titre officiel du document (dans le JO, dans le Recueil, dans les journaux officiels des EEMM, etc.)	TOUS SECTEURS Alphanumérique, texte avec indexation positionnelle
	VISA	-	Dispositions visées dans l'acte concerné	SECTEURS 1 à 7 Alphanumérique, texte avec indexation positionnelle
	CONSIDER	-	Considérations en fait ou en droit précédant le texte du document	SECTEURS 1 à 5 et 7 Alphanumérique, texte avec indexation positionnelle
	CORPUS	-	Partie textuelle proprement dite du document	SECTEURS 1 à 5, 7 et 9 Alphanumérique, texte avec indexation positionnelle
	FOOTNOTE	-	Note en bas de page du document	SECTEURS 1 à 5 et 7 Alphanumérique, texte avec indexation positionnelle
	ANNEXE	-	Annexes du document	SECTEURS 1 à 5 et 7 Alphanumérique, texte avec indexation positionnelle

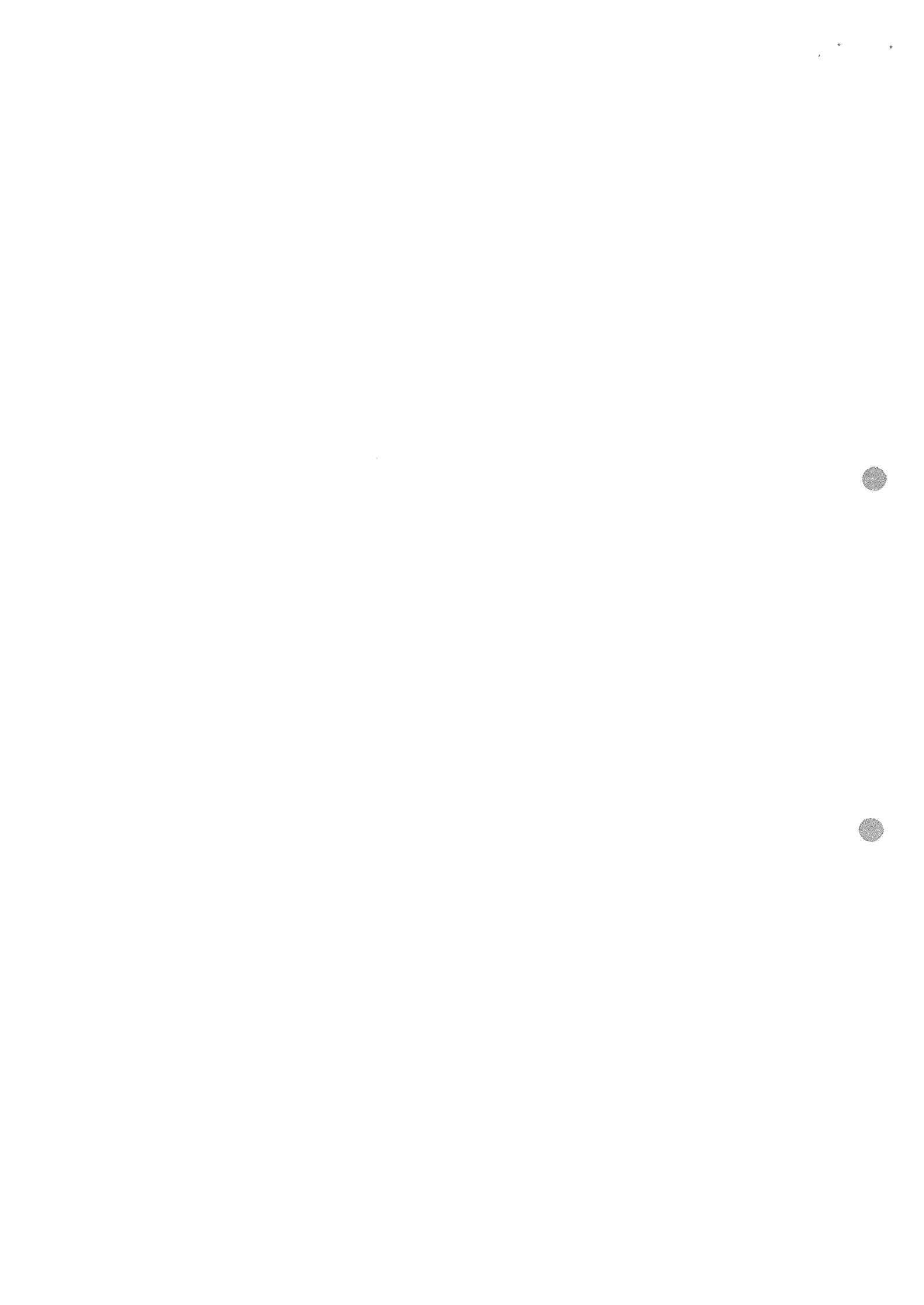


NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
TXT (suite)	RECTIF	-	Texte des rectificatifs de l'acte concerné	SECTEURS 1 à 5 Alphanumérique, texte avec indexation positionnelle
	SOMMAIRE	-	Résumé de l'arrêt tel que publié au Recueil de la jurisprudence	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	INDEX	-	Mots clés attribués à l'arrêt, tels que publiés Recueil	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	INTITULE	-	Nom des parties, objet du recours, composition de la Cour	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	MOTIFS	-	Texte de la partie en droit de l'arrêt	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	DEPENS	-	Texte de la décision de la Cour sur les dépenses	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	DISPOSI	-	Texte du dispositif de l'arrêt	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	CONCLUS	-	Texte des conclusions de l'avocat général	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle



SECTEURS DU NOUVEAU CELEX

1. TRAITES
2. RELATIONS EXTERIEURES
3. DROIT DERIVE ET DISPOSITIONS NATIONALES D'EXECUTION DES DIRECTIVES
4. DROIT COMPLEMENTAIRE
5. TRAVAUX PREPARATOIRES
6. JURISPRUDENCE DE LA COUR DE JUSTICE DE LA C.E.
7. VERSIONS DES TEXTES
8. JURISPRUDENCE NATIONALE RELATIVE AU DROIT COMMUNAUTAIRE
9. QUESTIONS PARLEMENTAIRES
10. DOCTRINE RELATIVE AU DROIT COMMUNAUTAIRE



LES GRANDS TRAITES DU COMMON COMMAND LANGUAGE (CCL)¹

Le CCL est un langage de commande standardisé qui facilite les recherches en ligne en mode conversationnel par l'emploi de noms, de commande significatifs et de leur abréviations d'une part, et d'autre part, par une syntaxe et une sémantique de langage simples dans leur utilisation.

1. Principales commandes

Le choix de la base de données s'opère par la commande BASE ou B;

Les données se visualisent par la commande SHOW ou S et s'impriment avec la commande PRINT ou P;

La commande DEFINE ou DEF permet de spécifier une édition ou un lexique par défaut;

L'historique de la recherche est affiché par la commande REVIEW ou R;

DELET ou DEL supprime une ou plusieurs étapes de recherche;

Les commandes UST et DISPLAY affichent l'environnement alphabétique et relationnel d'un terme;

BACK et MORE font remonter ou descendre l'écran;

Les commandes ? ou HELP assistent l'utilisateur et STOP arrête la session d'interrogation;

La commande FIND ou F, enfin, permet de rechercher directement les données pertinentes; elle est suivie par l'introduction d'éléments et d'opérateurs de recherche.

2. Syntaxe et opérateurs

Les opérateurs booléens sont AND, OR, NOT; les opérateurs de relations sont : = /. La virgule sépare les termes de recherche et les signes +, !, ?, = (opérateurs de distance) masquent un nombre différent de caractères.

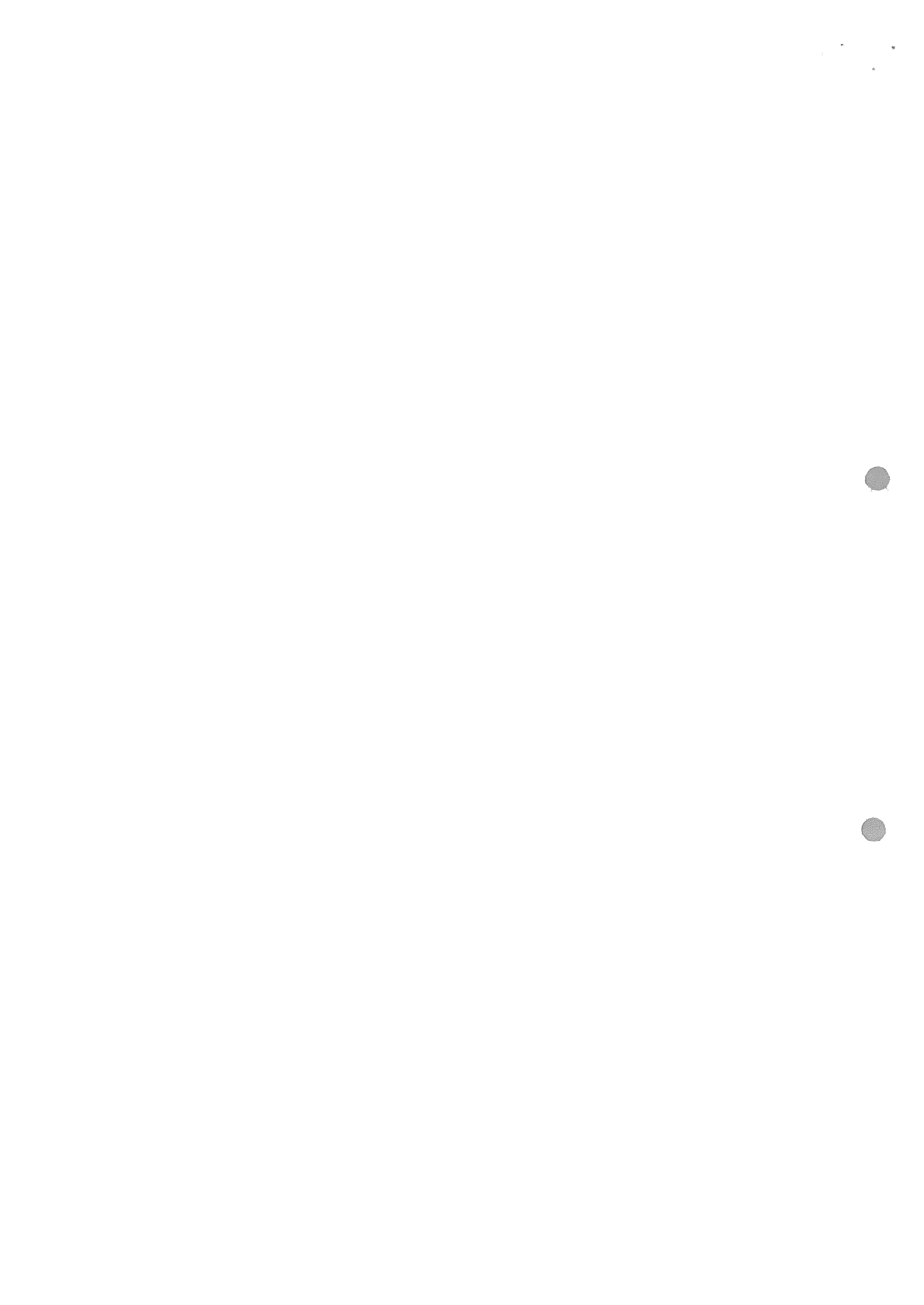
La syntaxe permet de combiner,² dans une même étape de recherche des recherches sur plusieurs champs² ainsi que différents opérateurs, sans ordre de priorité.

Pour limiter la recherche à un champ, il faut faire suivre le terme de recherche par un trait oblique et par le nom du champ.

Ex. : F CONSEIL/AUTEUR AND R/TYP.DOC

¹ Version ISO/DIS 8777

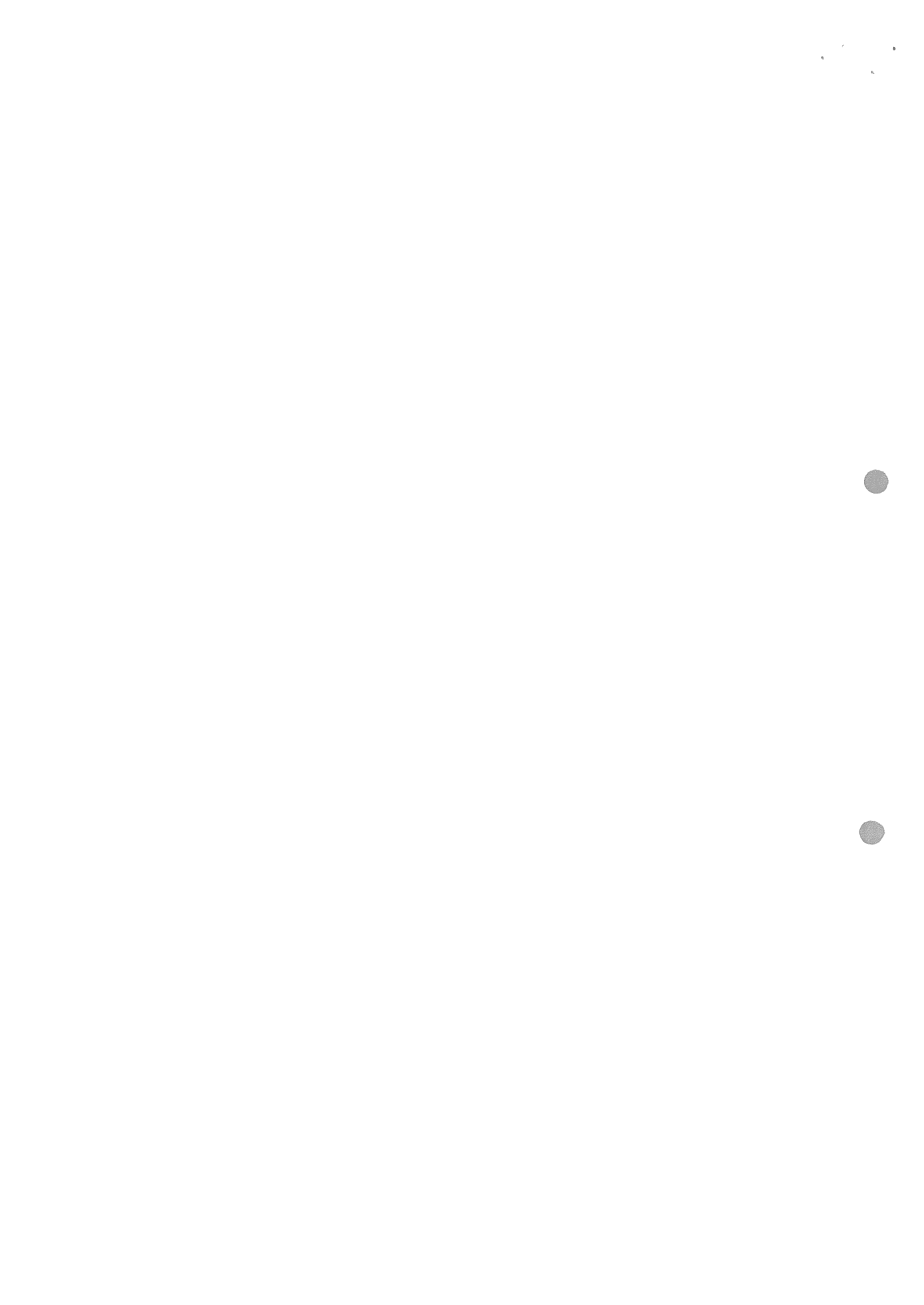
² Les noms des champs de base standardisés sont mnémoniques en anglais. Ex. : AU = auteur, DT = type de document etc.



Le terme ALL précédant le terme de recherche permet de travailler en mode transactionnel pour les termes multiples.

Ex. : F ALL 03+/REPERT

On trouvera une description in extenso des divers détails du langage commun de commande que le présent aperçu n'avait pas pour but de développer le dominant DRAFT INTERNATIONAL STANDARD ISO/DIS 8777.



VI. LES BENEFICES POUR L'UTILISATEUR FINAL

Le présent chapitre a pour objet de décrire de façon concrète, après un bref résumé des grands travaux du "project plan", les avantages que l'utilisateur final devrait logiquement obtenir du fait de la modernisation du système.

A. MISE EN PLACE DE LA NOUVELLE VERSION DE CELEX

Le projet global de développement est décomposé en deux sous-projets, la partie textuelle et la partie analytique. Les travaux de construction des deux sous-projets seront partiellement menés de front. Ainsi, le prototype "partie analytique" sera développé et testé pendant la construction du sous-projet 1 dans le but de pouvoir commencer la construction réelle du sous-projet 2 dès la mise en production de la partie textuelle.

Le project plan peut donc se résumer ainsi :

1. Sous-projet 1 : Production des données textuelles

- Construction du texterface	4	h/m
- Construction de la BDPT ¹	1,5	h/m
- Mise en place de l'interface BDPT - BDD ²	1,5	h/m
- Test et mise en place de BDPT	5	h/m
	<u>12</u>	<u>h/m</u>

2. Sous-projet 2 : Production des données analyse, diffusion et pilotage

- Construction de prototype	5,5	h/m
- Construction du module saisie/alimentation	4	h/m
- Construction de la BDPA ³	3,5	h/m
- Construction interface ₂ BDPA-BDD ²	6	h/m
- Construction de la BDD ²	4	h/m
- Test de mise en place de BDPA ³ et BDD ²	7	h/m
	<u>30</u>	<u>h/m</u>

Les ressources nécessaires au projet global se situeraient par conséquent près de 42 hommes/mois, tous investissements (SDA, SII, CC) confondus.

B. EVALUATION DE LA REDUCTION DES COÛTS D'UTILISATION

Sur base des coûts d'utilisation actuels d'une part et d'une évaluation des économies pouvant résulter d'une utilisation simplifiée du système issue de sa modernisation d'autre part, certaines hypothèses approximatives sont avancées ci-après avec les réserves d'usage. Une base d'évaluation plus scientifique de la réduction des coûts pourra être fournie dès achèvement des tests qui seront effectués sur un prototype.

1 Base de production des textes
2 Base de diffusion
3 Base de production de la partie analytique



1) Coûts de formation

Exemple : membre d'un cabinet juridique moyen situé à Frankfurt.

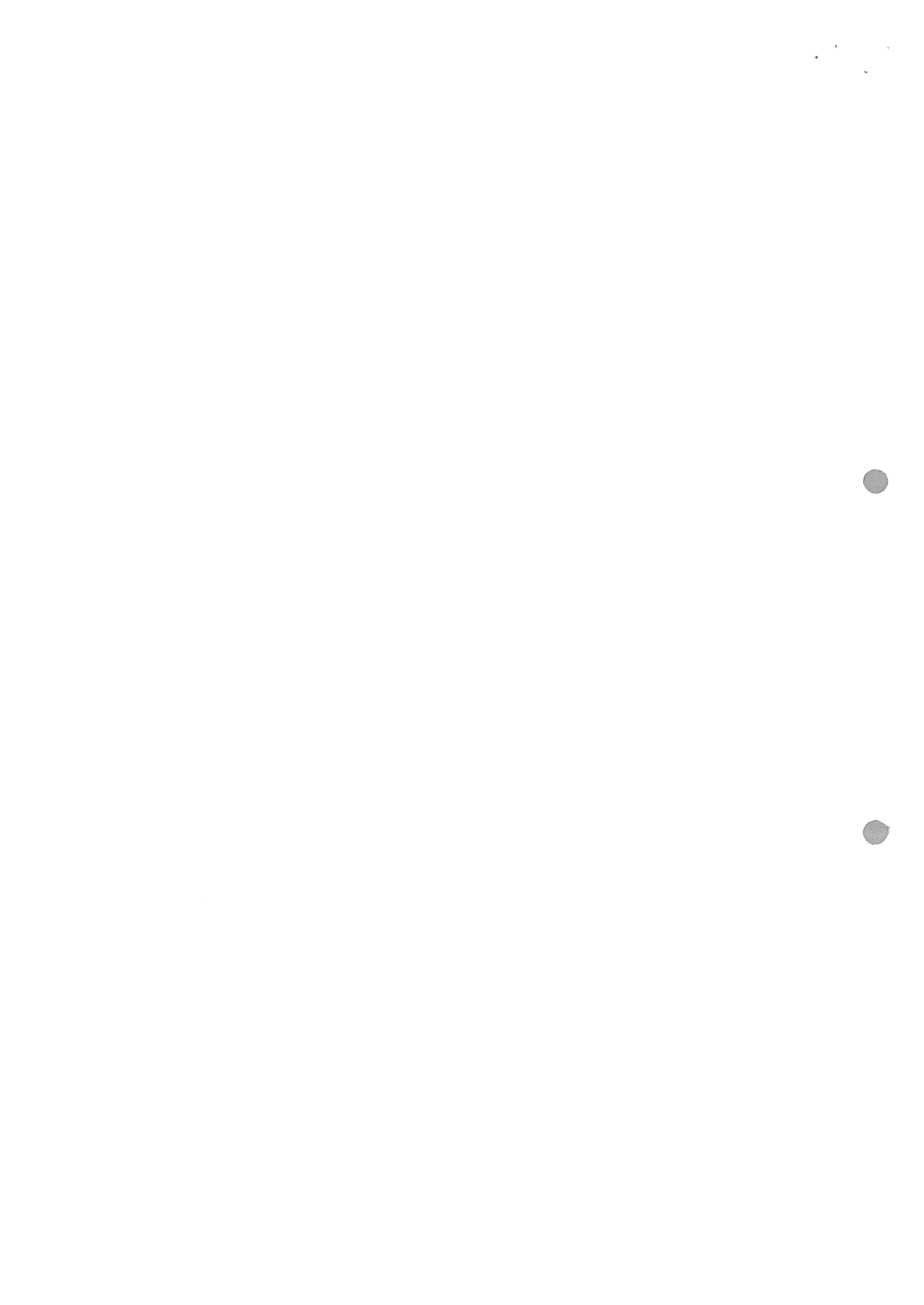
	<u>CELEX actuel</u>	<u>CELEX modernisé</u>	<u>BENE- FICE</u>
<u>a) Formation à Bruxelles</u>			
- Cours de formation initial à Bruxelles	(2 j)	(1 j) ⁴	
. droit d'inscription	4.000 FB	3.000 FB	
. frais de voyage et hôtel	+ 15.000 FB	11.000 FB	+ 30 %
. manque à gagner	+ 16.000 FB	6.000 FB	-
	<u>35.000 FB</u>	<u>20.000 FB</u>	
- Cours de perfectionnement	(1 j)	5	
. droit d'inscription	2.500 FB		
. frais de voyage et hôtel	11.000 FB		
. manque à gagner	8.000 FB		+ 100 %
	<u>21.000 FB</u>		
<u>b) Formation sur place</u>			
	(2 j)		
. droit d'inscription (pour 10 élèves)	3.000 FB +	6	
. manque à gagner	16.000 FB -		+ 65 % ⁶
	<u>19.000 FB</u>		

2. Coûts d'accès à l'information et l'exploitation des données

Les coûts de l'accès à l'information peuvent être chiffrés sur base du temps employé par l'utilisateur pour obtenir un résultat recherché en sachant que les prix de l'heure de connexion est de 4 200 FB et de l'heure de prestation de service de + 1 500 FB.

Dans l'hypothèse où un utilisateur occasionnel tel que décrit plus haut veut savoir quelles étaient par exemple, en 1983, les indications admises pour la désignation des vins de tables sur l'étiquetage (art. 3 du R/ 79/355), le scénario suivant peut être avancé :

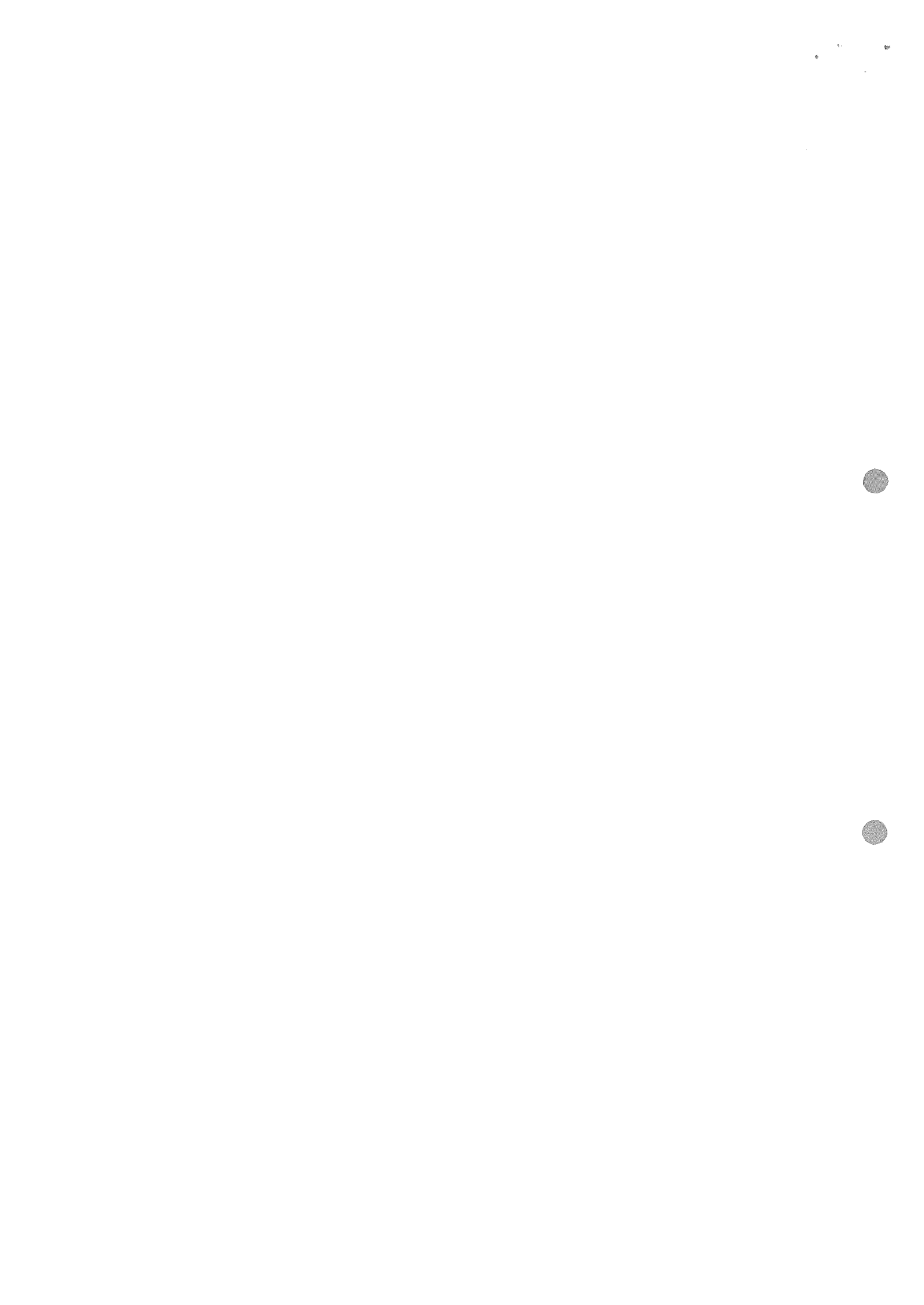
-
- 4 pourra être réduit à 1 jour en raison de la simplification de l'accès.
 5 pourra être supprimé en raison de la simplification de l'accès et de la disponibilité de cours sur support magnétique.
 6 pourra être supprimé en raison de la disponibilité du cours sur support magnétique.



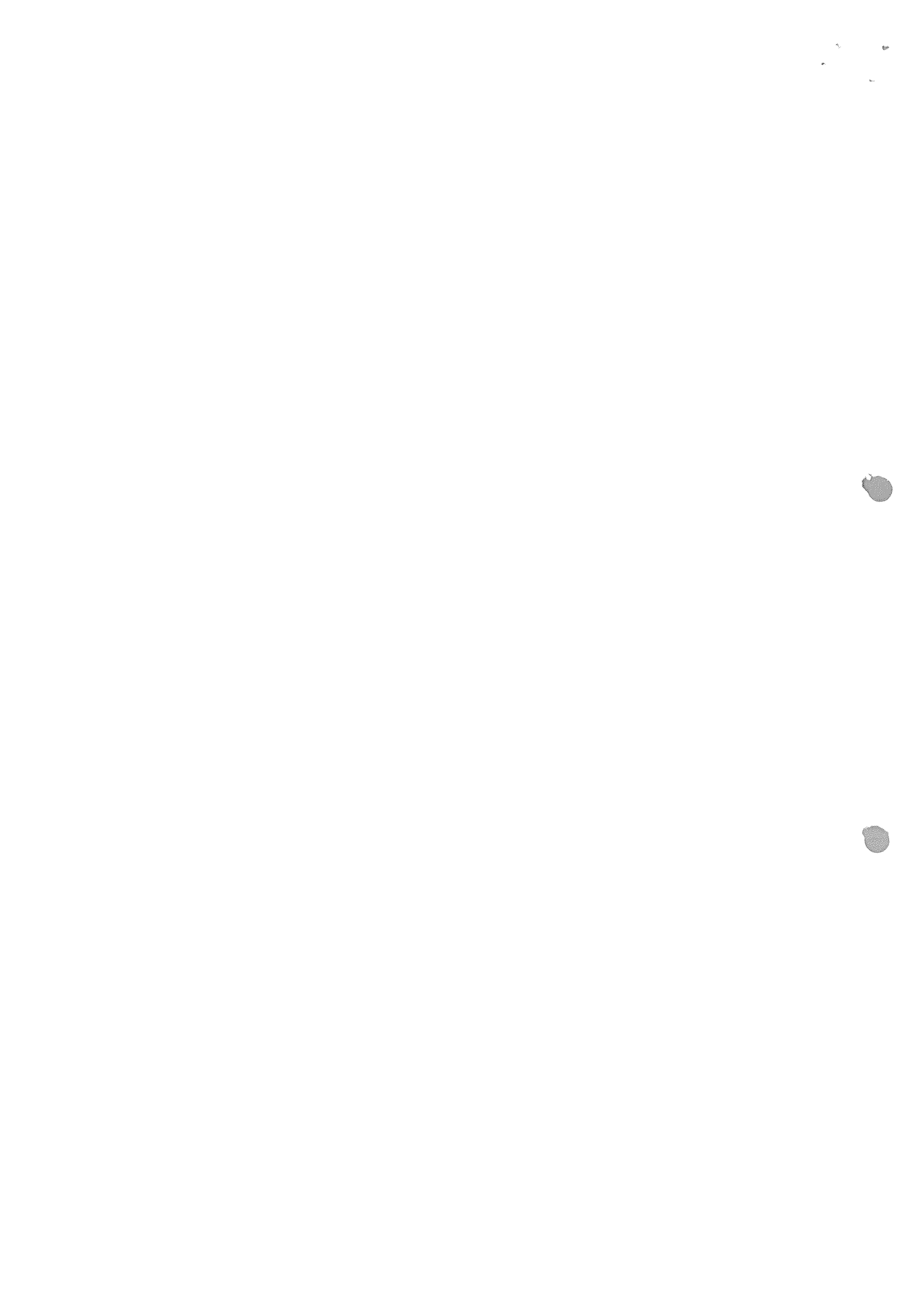
<u>PROCEDURE</u>	<u>TEMPS EMPLOYE</u>	<u>COUT DE CONNEXION</u>	<u>COUT DE PRESTATION DE SERVICE</u>
<u>CELEX actuel</u>			
- sélection et affichage de tous les documents "vins"	0,15 h	1 050 FB	375 FB
- sélection du sous-ensemble en vigueur en 1983	0,10 h	700 FB	250 FB
- sélection intellectuelle du seul texte pertinent par :			
. affichage et lecture des textes existants	0,50 h	3 500 FB	1 000 FB
. isolation des références de publication des textes manquants	0,15 h	1 050 FB	375 FB
. recherche des textes manquants dans JO et lecture	0,25 h	-	625 FB
- recherche des actes modificateurs des dispositions pertinentes de l'acte sélectionné	0,10 h	700 FB	250 FB
- reconstitution intellectuelle puis matérielle (copie, coupage, collage) des dispositions pertinentes et leur incorporation dans le document de travail	1,30 h	-	2 250 FB
TOTAL	3,25 h	7 000 FB	5 125 FB
COOT GENERAL			12 125 FB

<u>PROCEDURE</u>	<u>TEMPS EMPLOYE</u>	<u>COUT DE CONNEXION</u>	<u>COUT DE PRESTATION DE SERVICE</u>
<u>CELEX modernisé</u>			
- sélection et affichage des seuls documents pertinents	0,20 h	1 400 FB	500 FB
- sélection du seul document en vigueur en 1983	0,10 h	700 FB	250 FB
- affichage de la version modifiée du texte et print des dispositions pertinentes	0,10 h	700 FB	250 FB
- incorporation dans le document de travail	0,10 h	-	250 FB
TOTAL	0,50 h	2 800 FB	750 FB
COOT GENERAL			3 550 FB

Ces chiffres montrent clairement que les bénéfices que les utilisateurs pourraient retirer de la modernisation du système sont appréciables. Il faut cependant souligner à nouveau que l'évaluation du temps de recherche dans le système modernisé est théorique. Elle se base sur les déductions suivantes qui sont à confirmer par les tests du prototype :



- la durée de recherche devrait être réduite par la simplification de la structure et par la suppression, grâce au guidage, du tâtonnement de l'utilisateur (résultats partiels, stratégies de recherche inadéquates, etc.);
- l'existence d'une version codifiée et mise à jour des actes supprimerait l'examen long et fastidieux des nombreux actes modificateurs postérieurs;
- la couverture exhaustive en texte intégral éviterait un recours complémentaire au Journal officiel et la meilleure présentation des textes (typographie riche, formatage clair, etc.) permettrait d'incorporer les données directement dans un document de travail ou un dossier, sans qu'il soit nécessaire de re-dactylographier les données.



PROJET DE DEVELOPPEMENT ET MODERNISATION DE CELEX

ASPECTS UTILISATEURS

Janvier 1989



La modernisation de CELEX doit aboutir à une refonte en profondeur du système. Quelles que puissent être les améliorations attendues quant à la rationalisation de la gestion générale de l'application, il va de soi que la raison d'être du projet, son but principal, est de **fournir aux utilisateurs de CELEX la meilleure information possible en matière de droit communautaire.**

Cet objectif pourra être atteint par deux moyens, complémentaires l'un de l'autre :

- la **qualité des données** proprement dites;
- la **qualité des voies d'accès à ces données** qui seront mises à la disposition des utilisateurs.

Le premier de ces éléments sera essentiellement fondé sur l'amélioration des procédures de production des données, et relève des équipes responsables de l'administration du système au sein des institutions.

Par contre, la mise en oeuvre des moyens d'accès à l'information aura des répercussions directes sur les utilisateurs de CELEX. Il est donc indispensable que ceux-ci soient impliqués dès la phase de définition des bases de diffusion et des interfaces d'accès afin de garantir l'adéquation du projet à leurs besoins.

Le présent document, complémentaire de la préanalyse du projet de développement et de modernisation de CELEX, a pour objet de **présenter les aspects "utilisateurs" du projet**, en décrivant par quels moyens et quelles méthodes seront pris en compte les besoins de ceux-ci.

Après un bref **rappel des objectifs généraux** de la modernisation de l'application, il sera procédé à :

- un résumé des **améliorations prévues en matière de gestion générale**, avec les conséquences non négligeables qui en découleront pour l'utilisateur final de CELEX;
- une **définition des différentes catégories d'utilisateurs** et de leurs besoins spécifiques;
- une **description des différents moyens d'accès à l'information** qui seront mis à la disposition des utilisateurs;
- la **proposition d'une méthode de communication** à établir entre les responsables du projet et les utilisateurs pour que ceux-ci puissent s'exprimer sur les options envisagées;
- un **chiffrage des bénéfices** apportés par l'utilisation du nouveau système.

I. RAPPEL DES OBJECTIFS

Les principaux objectifs de l'actuel plan de développement et de modernisation de CELEX ont été exposés dans le "problem statement" présenté au Groupe Informatique juridique du Conseil lors de sa réunion des 11 et 12 juin 1987.

Ils consistent en :

A. La menée à terme des objectifs du plan précédent (1983 - 1985).

Ceci comprend :

- la mise en place de la version danoise (novembre 1988);
- la mise en place de la version grecque (courant 1989);
- la création des versions espagnole et portugaise (en fonction de la disponibilité des ressources);
- la création du secteur 8 (jurisprudence nationale se référant au droit communautaire);
- la création du secteur 10 (doctrine publiée relative au droit communautaire).

B. L'amélioration de l'alimentation, de la mise à jour et de la gestion générale du système.

C. L'amélioration de la convivialité d'accès à l'information, laquelle doit se fonder sur :

- l'enrichissement du contenu de la base, aussi bien en ce qui concerne les données que leur degré d'actualité;
- la simplification de la structure logique de l'application;
- une présentation physique des données permettant leur utilisation immédiate;

- la mise à disposition des utilisateurs d'interfaces conviviales d'accès à l'information, adaptées à leur degré d'expertise en matière juridique et/ou documentaire

Il va de soi que la réalisation de ces objectifs, dans le contexte d'un projet d'une telle ampleur, sera fonction de l'allocation des ressources nécessaires.

II. EFFETS POUR L'UTILISATEUR D'UNE GESTION AMELIOREE

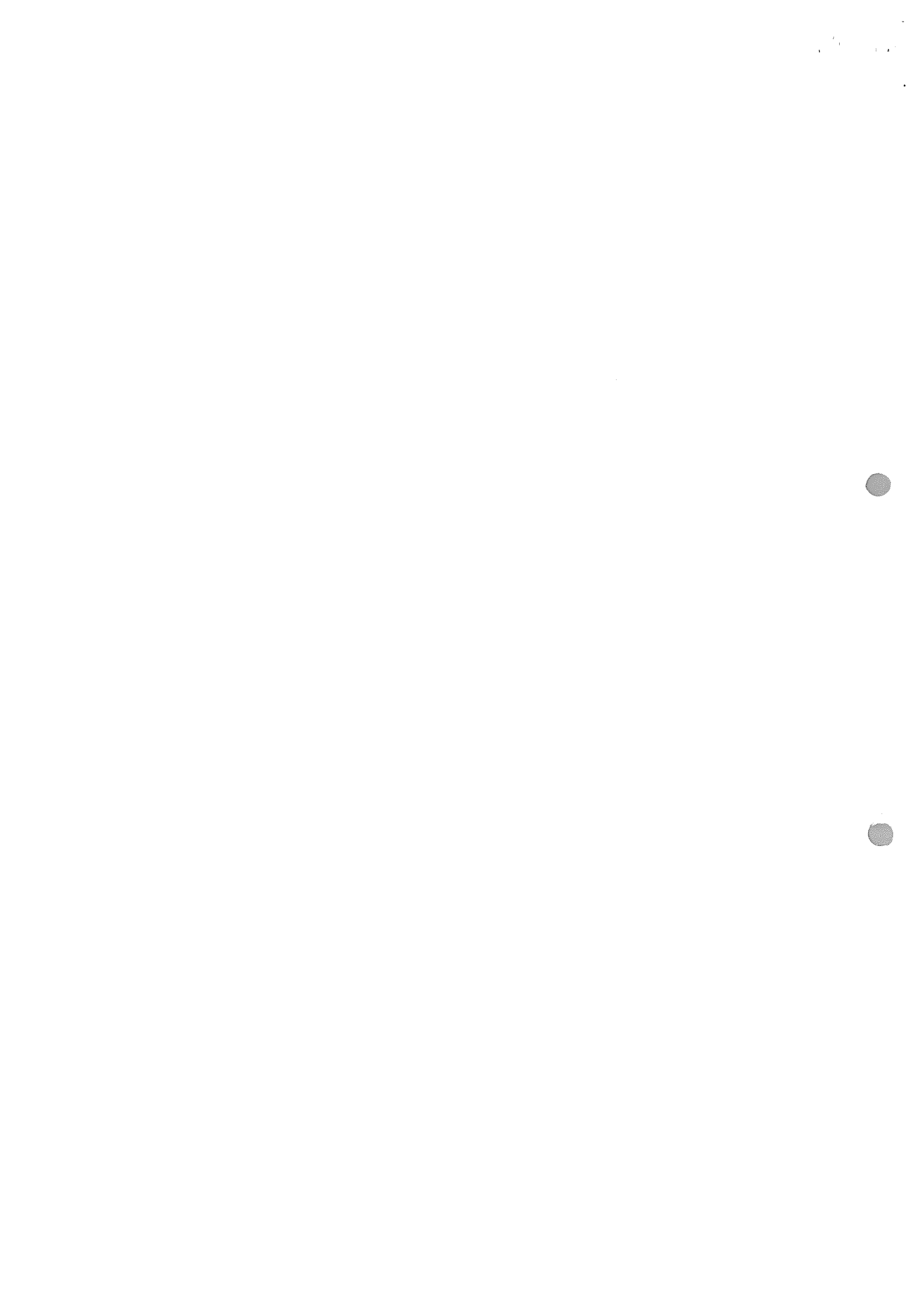
L'optimisation des processus d'alimentation, de mise à jour et de contrôle de l'application sera obtenue par :

- la simplification de la structure de la base de production des données;
- la création d'un instrument d'analyse documentaire en ligne, avec contrôle formel et logique des données introduites;
- la rationalisation des traitements des données et des chaînes d'alimentation de la base de production, puis, à partir de celle-ci, des bases de diffusion;
- la mise en place d'une interface de chargement des textes en format normalisé;
- la modularisation de l'architecture du système, permettant une adaptation souple et peu coûteuse à de nouveaux besoins ou à l'évolution technologique.

L'accroissement des performances et de la productivité générale du système permettra d'améliorer la qualité du service fourni aux utilisateurs en apportant une réponse positive à des besoins déjà exprimés en matière de :

- mise en place des versions linguistiques manquantes;
- mise en place des derniers secteurs à créer;
- parachèvement du chargement des textes intégraux;
- fiabilité encore accrue des données;
- enfin, et surtout, meilleure actualité de l'information.

Pour ce dernier point, auxquels les utilisateurs sont à juste titre très sensibles, l'objectif poursuivi serait la disponibilité à l'écran de l'information dans un délai d'une semaine après la publication.



III. DETERMINATION DES CATEGORIES D'UTILISATEURS ET DE LEURS BESOINS

Afin de disposer d'un maximum de données factuelles sur les conditions d'emploi des bases de données juridiques en Europe, sur les besoins exprimés par les utilisateurs et sur les moyens d'y apporter une réponse satisfaisante, la Commission a fait procéder à une étude générale du marché.

Les auteurs de cette étude ("General acces to legal databases/systems in Europe, among others those produced by public authorities in Member States and by Community Institutions") ont examiné les principales bases de données juridiques en Europe, et notamment CELEX, et ont défini, sur base de nombreuses enquêtes, les différentes catégories d'utilisateurs du système. Par appartenance professionnelle, ceux-ci se répartissent en :

- Institutions européennes;
- Parlements nationaux;
- Administrations;
- Juridictions nationales;
- Praticiens privés du droit;
- Services juridiques d'entreprises;
- Universités;
- Bibliothèques;
- Centres de documentation et courtiers d'information.

A l'intérieur de ces grandes catégories, il ressort des enquêtes effectuées qu'il est fait, pour 50 % de l'utilisation totale, appel à des "utilisateurs intermédiaires", chargés de rechercher dans la ou les bases, les informations nécessaires aux travaux de l'utilisateur "réellement final".

Les raisons du recours à ces utilisateurs intermédiaires sont de plusieurs ordres :

- le consommateur final n'est pas formé à l'interrogation;
- le consommateur final n'a pas le temps d'effectuer lui-même les recherches;
- l'utilisateur final éprouve une réticence à l'emploi d'un clavier et d'un terminal;
- les recherches effectuées par l'utilisateur final lui-même sont plus coûteuses ("le temps d'un juriste est précieux");
- les utilisateurs intermédiaires sont plus expérimentés et fournissent des résultats plus pertinents.

Selon l'étude, les utilisateurs intermédiaires se recrutent parmi :

- des courtiers d'information indépendants;
- des courtiers d'information dépendant du fournisseur de celle-ci (p.ex. "help desk" rattaché à un serveur);
- des "information officers" spécialisés dans l'interrogation des bases;
- des membres de l'équipe responsable de l'administration de la base;
- des bibliothécaires ou documentalistes;
- des juristes débutants ou stagiaires à qui les travaux de recherche ont été délégués;
- des juristes professionnels particulièrement experts dans l'utilisation des bases juridiques.

En fait, et quelle que soit leur appartenance professionnelle, il faut considérer l'existence de **trois types d'utilisateurs interrogeant la base**, il s'agit des utilisateurs :

- occasionnels;
- expérimentés;
- professionnels.

Il convient, à ce stade, de rappeler la **vocation juridique** de CELEX, et donc les besoins spécifiques auxquels cette application doit répondre en premier lieu. On ne peut mettre sur le même plan un outil comme CELEX et une base bibliographique générale, comme on ne peut mettre sur le même plan un journal quotidien d'information et la "Revue trimestrielle de droit européen".

Ceci n'exclut pas cependant que, par la mise en place de technologies appropriées, l'accès à l'information sur le droit communautaire soit facilité pour un public de plus en plus large.

Les conclusions de l'étude, confortées par celles tirées par les responsables de CELEX dans l'exécution de leur tâches de production des données, de relations interinstitutionnelles, de formation des utilisateurs internes et externes et de suivi de ces utilisateurs, confirment qu'il est fait recours à la base de données pour répondre à des besoins précis. Parmi les recherches typiques, on peut citer :

- vérification de l'état d'un acte juridique déterminé (adoption, dernière modification, etc ...);
- vérification d'un ensemble de références et de citations;
- utilisation de la base comme contrôle d'exhaustivité d'une recherche "manuelle";
- recherche de précédents ou de lignes de jurisprudence;
- identification de tout le matériel pertinent avant d'entamer une affaire (constitution d'un dossier);

En ce qui concerne les stratégies de recherche utilisées, il se révèle qu'en général, les utilisateurs occasionnels ne recourent qu'à un sous-ensemble relativement réduit des possibilités offertes par le système, les recherches plus sophistiquées étant le plus souvent effectuées par des utilisateurs expérimentés.

Par ailleurs, les recherches très générales (du type "tout sur ...") restent minoritaires par rapport à celles concernant un point de droit précis, effectuées sur des critères bien spécifiques comme des citations, des dates, des références croisées, etc ...

L'étude Butler Cox précitée envisage un certain nombre de mesures permettant une meilleure utilisation des bases de données juridiques, comme par exemple :

- améliorer la logique interne de l'application, afin que les recherches reçoivent une réponse cohérente et complète, indépendamment du secteur ou du type d'acte concerné;
- enlever du langage de commande des "détails irritants" tels que codes, abréviations, etc ...
- créer des "interfaces utilisateur" flexibles;
- offrir la possibilité d'utiliser des terminaux spécialement adaptés, permettant un meilleur accès et une meilleure restitution de l'information;
- améliorer les systèmes de diagnostic et d'assistance (help);
- considérer la formation, la documentation, le support et l'interface utilisateur comme un ensemble interactif dont tous les éléments, ensemble, doivent rendre l'utilisation du système plus souple et plus performante.

Telles que résultant des enquêtes effectuées, ces propositions concernent les bases de données juridiques en général, et pas seulement CELEX; elles ont constitué cependant une base de réflexion importante dans l'analyse du projet de modernisation.

La réponse qui sera donnée aux besoins exprimés par les utilisateurs sera déterminante pour l'avenir de CELEX, c'est donc elle qui conditionnera les options stratégiques de développement du nouveau système, particulièrement en ce qui concerne les moyens d'accès à l'information.

IV. LES VOIES D'ACCES A L'INFORMATION CONTENUE DANS CELEX

Les problèmes décrits sous ce point, et les solutions envisagées se rapportent à la base de diffusion de référence, telle que diffusée par les Institutions sur le Centre de Calcul de Luxembourg.

Les serveurs externes restent libres de mettre en place les solutions qu'ils estiment les plus appropriées à leur environnement technique et de clientèle.

Les difficultés rencontrées dans l'utilisation de CELEX dans sa version actuelle sont dûes à :

1. des lacunes dans le contenu de la base (absence de certains textes, délai de mise à jour, ...);
2. une structure difficile à assimiler et à mémoriser (nombre trop élevé de rubriques, différences parfois injustifiées entre secteurs couverts, etc ...);
3. un manque de convivialité du logiciel d'interrogation;
4. un matériel pédagogique peu commode à utiliser et dont la mise à jour pose des problèmes.

L'amélioration de la gestion, avec ses conséquences telles que décrites sous le point II ci-dessus, supprimera la première de ces causes de désaffection à l'égard de CELEX. Cependant, pour complet que puisse être le contenu d'une application, sa valeur aux yeux des utilisateurs sera directement proportionnelle aux possibilités offertes à ces derniers d'accéder aux données et de les exploiter. C'est pourquoi, dans le cadre du projet de modernisation, un effort particulier sera consacré :

- à la structure de l'application;
- aux logiciels et interfaces d'interrogation;
- à l'ensemble des moyens pédagogiques;

de manière à ce que les utilisateurs de CELEX puissent bénéficier des possibilités offertes par la base, lesquelles ne seront **en aucun cas** amoindries par une simplification abusive mais au contraire rendues plus directement exploitables par les praticiens du droit et de l'information juridique pour lesquels l'application a été créée et aux besoins desquels elle doit correspondre.

A. STRUCTURE

La complexité de la structure actuelle de CELEX se traduit d'abord pour ses utilisateurs par une difficulté d'assimilation et de mémorisation de tous les paramètres. Une première étape dans la voie de la rationalisation consisterait à adapter ces paramètres (noms des lexiques, des champs, etc ...) aux normes et règles le plus généralement admises ou pour lesquelles un processus de standardisation est en cours dans le domaine des bases de données documentaires. D'autre part, il existe une relation directe entre la complexité de la structure, telle qu'elle est perçue par les utilisateurs, et la disposition des données dans la base.

La performance de l'accès à l'information est tributaire de la présentation de cette dernière, dont l'amélioration doit tenir compte de l'expérience acquise dans l'interrogation des bases documentaires, des besoins exprimés par les utilisateurs et des nouvelles possibilités de la technique. Les données formatées devront être à la fois plus claires et plus cohérentes alors que les textes devront offrir à la fois une meilleure lisibilité et un meilleur taux de précision et de rappel lors de l'interrogation.

De fait, de par sa vocation de base de données de type "texte intégral", on pourrait concevoir que CELEX ne comporte qu'un nombre réduit de rubriques analytiques. Toute l'information juridique relative à un acte est en principe présentée dans le corps de celui-ci tel qu'il est publié, et devrait donc pouvoir être recherchée et extraite par interrogation sur le texte intégral du document. Cependant, cette information n'est pour ainsi dire pas structurée, et ne le sera pas à terme prévisible. De plus, outre les risques de "silence" en cas d'absence d'un texte, certains éléments d'information tels que la classification par mots-clés¹, le nom du juge rapporteur, la date explicite de

¹ Il est envisagé de généraliser l'implantation du thésaurus EUROVOC 2 dans CELEX si l'expérience d'introduction progressive dans le champ DESCRIPT, actuellement en cours au Parlement européen, est favorablement accueillie par les utilisateurs.

fin de validité, etc ... n'y figurent pas et leur introduction dans les rubriques bibliographiques est le résultat du travail d'analyse effectué au sein des différentes équipes chargées de l'alimentation de CELEX dans les Institutions.

Pour offrir le maximum de possibilités de recherche, la base de données a été conçue dès l'origine pour contenir une partie textuelle et une partie analytique, cette dernière devant faciliter l'accès à l'information via l'emploi d'un vocabulaire contrôlé. La redondance d'information entre les deux parties est dès lors inévitable, mais volontaire. Ce qui est par contre inacceptable est que la partie analytique comporte une part non négligeable de redondance sur elle même.

Ce défaut sera pallié par le regroupement des données indexées au sein de quatre fichiers inverses dans lesquels les informations seront réparties en fonction de leur nature (données bibliographiques, dates, relations entre actes juridiques et textes) et dont la structure permettra de prendre en compte, sous un format standardisé aux maximum, les données relatives à chaque sorte de document relevant de la couverture de CELEX.

Un tableau des nouveaux champs et lexiques figure en annexe 1.

B. LOGICIELS ET INTERFACES D'INTERROGATION

L'utilisateur de CELEX aura le choix entre trois moyens d'accès au contenu de la base :

- le logiciel de recherche documentaire dans son mode natif;
- le Common Command Language (CCL);
- une interface conviviale d'accès guidé par menus (User Friendly Interface - UFI).

1. Le logiciel documentaire

A l'heure actuelle, CELEX est interrogeable dans sa version de référence sur le Centre de Calcul de la Commission via le logiciel documentaire MISTRAL V5 de BULL.

En ce qui concerne la nouvelle base de diffusion dans ses différentes versions linguistiques, l'étude de faisabilité sera effectuée dans l'hypothèse du maintien de ce logiciel comme moyen d'interrogation, étant bien entendu :

- qu'un effort particulier sera consacré à éviter tout risque de captivité de l'application, notamment en ce qui concerne la structure;
- que la ou les interfaces conviviales ne seront en aucun cas liées à ce logiciel;
- que l'équipe projet suivra activement les progrès de tout autre produit de recherche documentaire pouvant être utilisé, notamment BASIS.

2. Le Common Command Language (CCL)

L'existence d'un grand nombre de systèmes et logiciels de recherche documentaire pose un problème à l'utilisateur, confronté à des langages et types de commandes trop variés. L'utilisation de ces systèmes sera facilitée par la mise en place d'un lot normalisé de commandes de base, avec leurs règles d'utilisation.

Le standard international de CCL (norme ISO TC46 - actuellement au stade de Draft International Standard (DIS 8777) spécifie une liste de commandes permettant la recherche documentaire en mode interactif.

Cette norme n'apportera ni restriction ni obstacle à l'emploi éventuel du logiciel documentaire dans son mode natif (non standardisé) et restera compatible avec la mise en place d'une interface conviviale.

Une brève description du CCL figure en annexe 2.

3. L'interface conviviale (User Friendly Interface)

L'interface conviviale doit, à terme, devenir le chemin d'accès à CELEX le plus couramment utilisé, et ceci pour tous les types d'interrogation effectuées sur la base. S'il est inévitable que la complexité du droit communautaire se reflète dans la complexité du contenu de la base, le rôle de l'interface sera de **proposer à l'utilisateur un choix d'options suffisamment clair et complet** pour que, **quel que soit son niveau d'expertise**, il obtienne vite et bien l'information qui lui est nécessaire. Cette approche exclut volontairement le principe selon lequel une recherche "fine" implique la connaissance préalable de commandes complexes et de leur syntaxe d'utilisation : l'interface doit permettre, bien sûr, de poser des questions simples ou très générales, mais aussi (et surtout) répondre aux besoins d'un juriste confronté à une recherche sophistiquée.

Les spécifications fonctionnelles de l'interface conviviale seront, d'un point de vue général, les suivantes :

- système de guidage par menus, avec possibilité, à tout moment d'obtenir des écrans d'assistance;
- présentation en mode écran;
- possibilité, si désiré, de court-circuiter l'arborescence;
- lorsqu'il est nécessaire que l'utilisateur introduise un paramètre (mot-clé, valeur, etc ...), la consultation de la liste des termes doit être possible;
- recours limité aux opérateurs de logique booléenne, mais possibilité de les employer si désiré (tels quels ou traduits en langage naturel) notamment pour combiner les résultats des différentes étapes;
- adaptation des écrans de visualisation en fonction de la recherche effectuée (par exemple la visualisation d'actes recherchés par un critère de dates doit faire apparaître celles-ci bien en évidence par rapport aux autres données);

Sur la base de ces spécifications générales, la conception des menus proposés aux utilisateurs de CELEX sera fondée :

- d'une part, sur le principe d'un cheminement logique de recherche, partant de critères de sélection génériques (par exemple les secteurs, l'année de publication, etc ...) et s'affinant progressivement par combinaison de ces critères ou par recours à des critères plus spécifiques, propres au type d'acte plus spécialement recherché;
- d'autre part sur la possibilité de sélectionner directement un ou des actes connus, sur lesquels une information précise est recherchée (un grand nombre de questions posées sont du type "dans quel JO a été publié tel acte", "quelles sont les modifications apportées au règlement ...", "quels est le numéro de la proposition de la directive ..." etc ...).

L'**assistance** à l'utilisateur en cours de recherche devra comprendre des écrans explicatifs visualisables sans interruption de la stratégie et décrivant, selon les cas :

- la syntaxe des commandes (lors de l'emploi du CCL ou du logiciel documentaire en mode natif),

- la logique d'une question appelée par le menu, ainsi que le format selon lequel l'utilisateur doit entrer d'éventuels paramètres (par exemple "la requête ... vous permet de sélectionner les actes en vigueur à une date donnée. Elle s'applique aux documents des secteurs 2, 3 et 4 (législation). La date doit être introduite sous la forme AAAA/MM/JJ...");
- le contenu des rubriques (par exemple "La rubrique TRIB.NAT fait référence aux décisions de juridictions nationales rendues dans le cadre d'un renvoi préjudiciel. Les codes A ou P indiquent si la décision est antérieure ou postérieure à l'arrêt de la Cour ...").

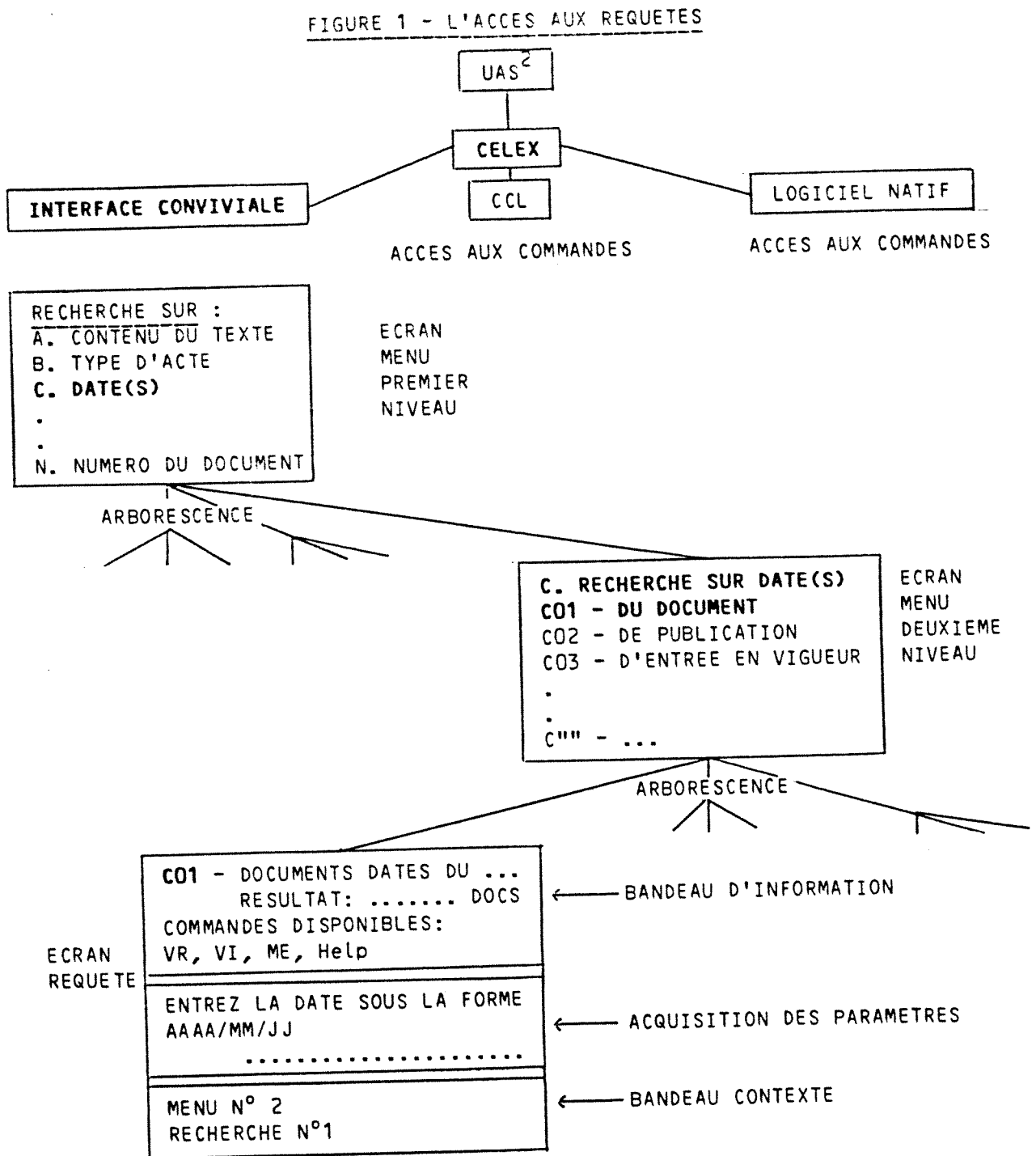
La conception d'une interface réellement efficace et conviviale permettra d'éviter l'erreur selon laquelle seul l'utilisateur chevronné du langage documentaire a la possibilité de poser une question complexe.

La structuration des stratégies (identifiées par des mnémoniques) en **menus arborescents** permet à l'utilisateur d'adapter sa recherche à sa connaissance du sujet (et non du langage ou de la structure de la base). A cette fin les menus proposés seront peu nombreux (4 à 6) tout en présentant chacun un grand choix de stratégies de recherche. Par ailleurs, la possibilité de sélection directe d'une stratégie par son mnémonique figurant dans le menu favorise à la fois l'apprentissage rapide de l'utilisateur débutant et l'interrogation performante de l'utilisateur plus expérimenté.

En cours de recherche, l'ergonomie au sens large du terme revêt une importance particulière tant quant à la présentation des écrans (standardisation, stabilité) qu'à leur contenu (information, clarté du dialogue, affichage du contexte dans lequel se situent les deux éléments précédents). Tous les éléments nécessaires au choix d'une commande sont visibles. Un écran d'assistance est toujours accessible.

Enfin, **l'affichage des documents sélectionnés** doit autoriser plusieurs modes de visualisation, celle-ci constituant en fin de compte le véritable accès final à l'information contenue dans la base.

La logique de l'interface conviviale CELEX peut être représentée de la manière suivante :



² USER AGENT SERVICE dont l'objectif général est l'amélioration de l'accès aux services informatiques pour la création d'outils tels que user agent (guide des services), interfaces conviviaux de recherche (UFI), système de type

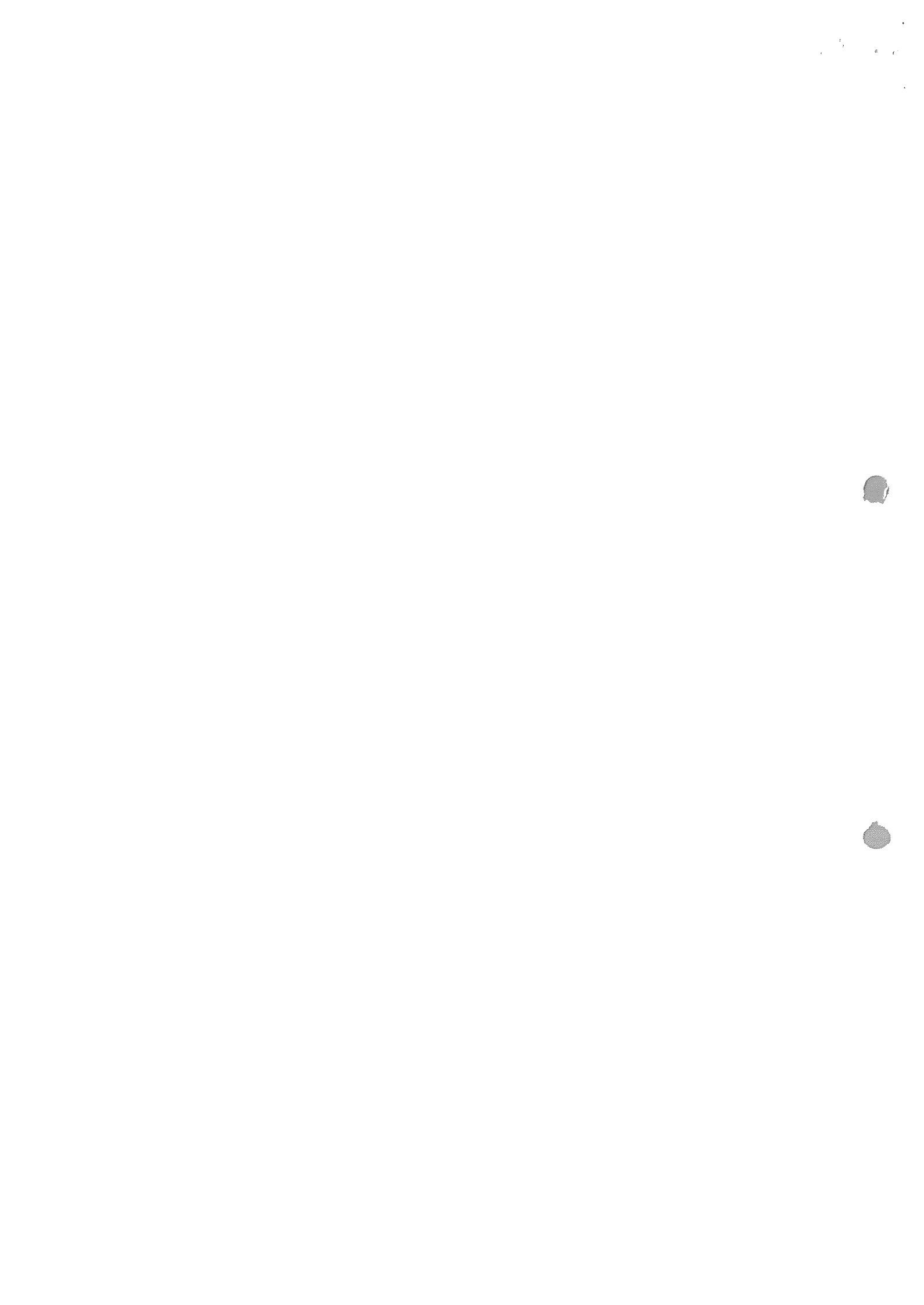
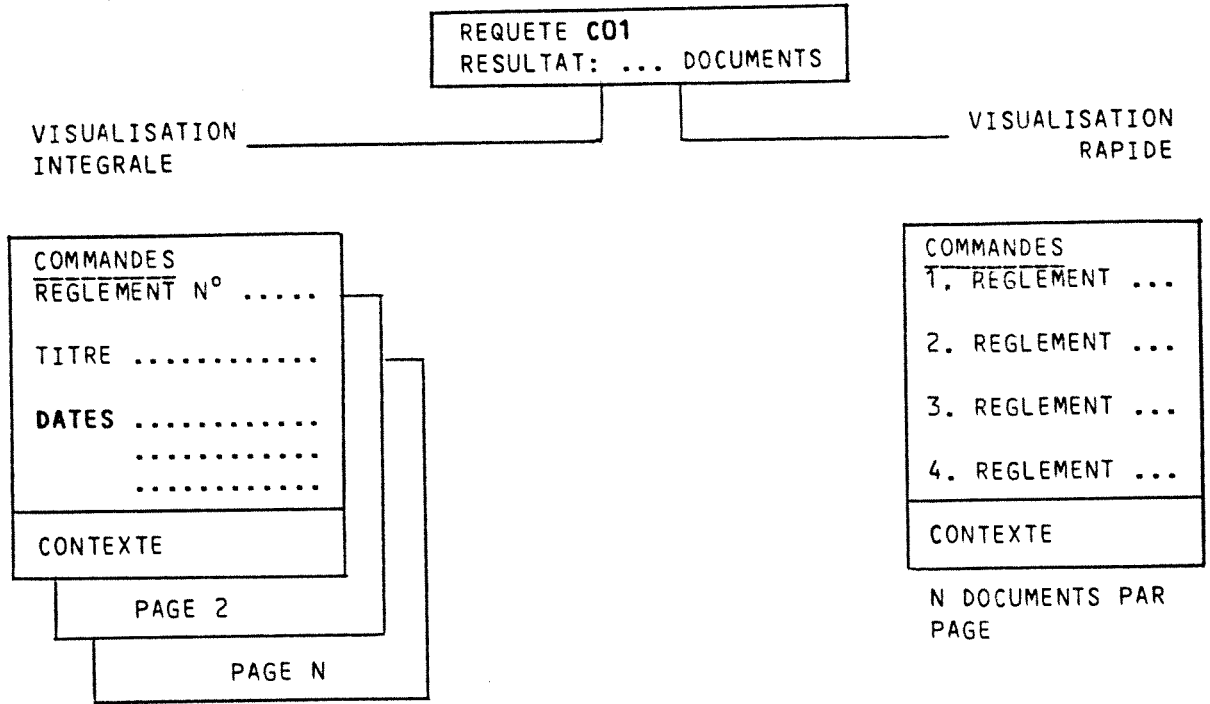


FIGURE 2 - L'AFFICHAGE DES DOCUMENTS

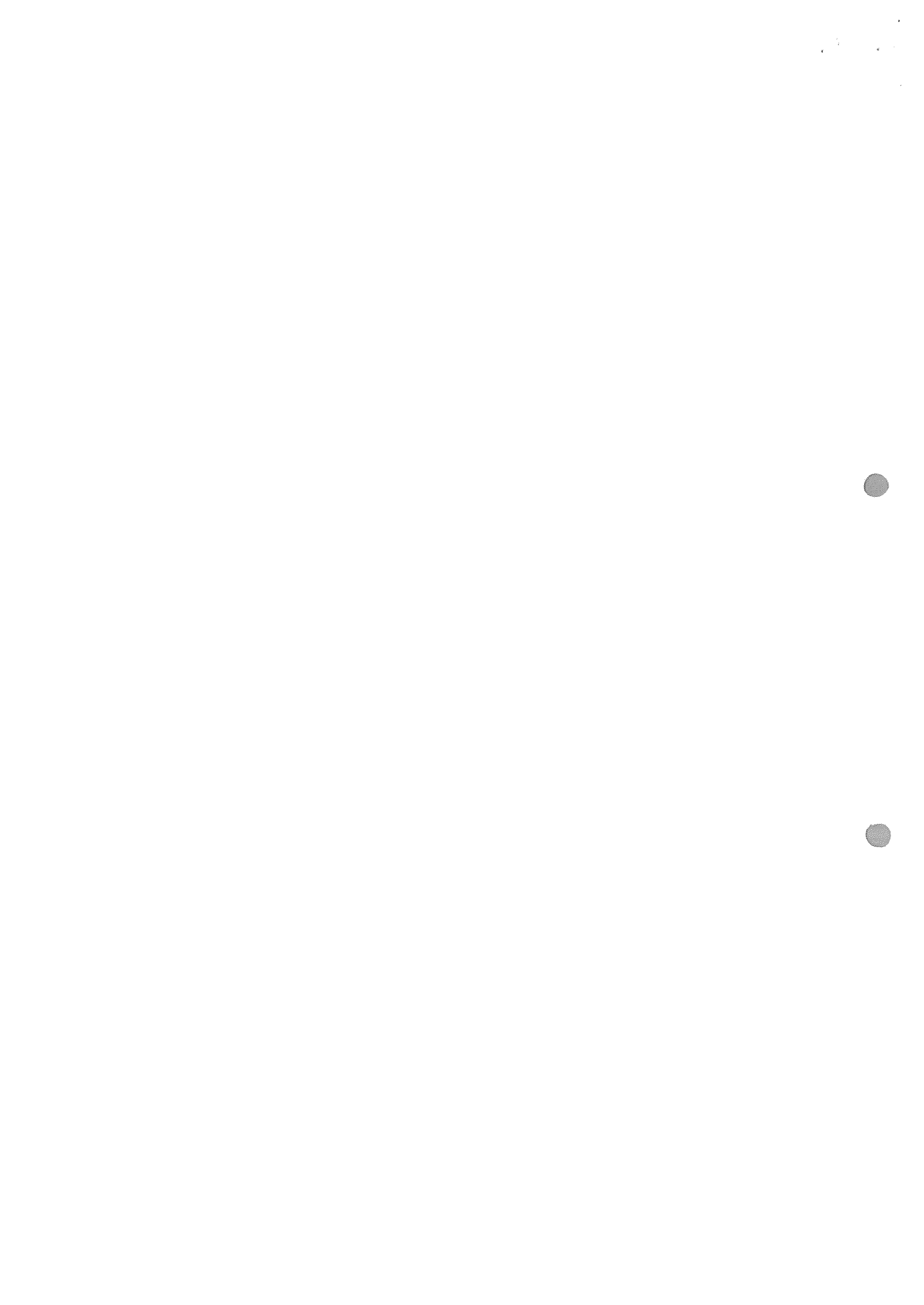
La restitution à l'écran des informations sélectionnées s'organise autour de deux éléments : les **commandes** de visualisation et la **présentation** des documents. Les commandes, outre les possibilités normales d'avance et de recul sur page ou document doivent comprendre une visualisation "rapide" et une visualisation "intégrale".



En fonction de la question posée, la présentation du document, structuré en "pages logiques", fait apparaître en premier lieu les éléments jugés les plus pertinents.

Dans le cas d'une structuration complexe, la première page peut contenir un "sommaire" du document ("les actes modifiés figurent en page 3 ...").

L'extension des possibilités d'assistance immédiate à l'utilisateur s'inscrira de façon complémentaire à l'ensemble des moyens pédagogiques mis à sa disposition.



V. PROPOSITION POUR UNE METHODE DE COMMUNICATION

Pour permettre aux utilisateurs de CELEX de se prononcer valablement sur les options envisagées quant à la modernisation du système, il importe que ces options leur soient présentées sous la forme la plus claire et pratique possible.

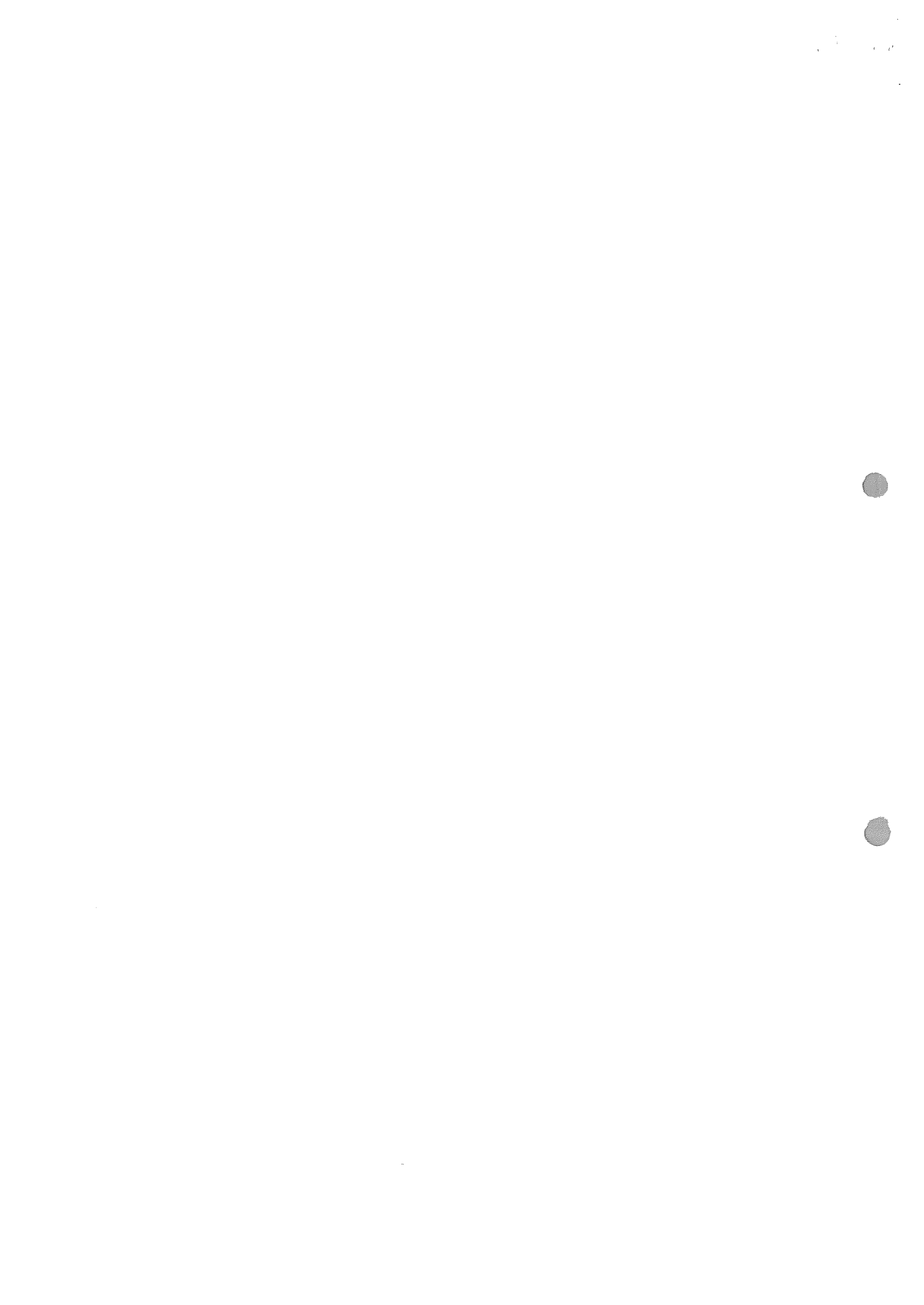
Cette information ne pourrait être fournie de façon plus concrète que par la mise en place et la présentation d'un **prototype** de la nouvelle version de CELEX. Ce prototype, contenant un nombre raisonnable de documents significatifs, sera rendu accessible et permettra de vérifier l'adéquation du projet aux besoins des utilisateurs.

Afin de coordonner la transmission de l'information entre les intervenants, il sera procédé à l'organisation de deux **cercles d'utilisateurs** (interne et externe aux Institutions) auxquels sera fourni un accès au prototype et un projet de manuel utilisateur.

Après la fixation d'une période de test, chacun des groupes fera part aux responsables du projet de ses remarques, aussi bien sur la structure et la présentation des données que sur les interfaces d'interrogation et le matériel pédagogique envisagés.

Une fois le projet de modernisation concrétisé, le prototype pourrait être conservé dans sa version définitive et mis à la disposition des utilisateurs potentiels de CELEX pour leur permettre de mieux comprendre les possibilités du système avant de souscrire à un contrat d'abonnement.

Par ailleurs, la pratique de tests exhaustifs sur une base modèle permettra un chiffrage précis des bénéfices apportés par l'utilisation de la nouvelle version de CELEX par rapport à sa version actuelle.



VI. LES BENEFICES POUR L'UTILISATEUR FINAL

Le présent chapitre a pour objet de décrire de façon concrète, après un bref résumé des grands travaux du "project plan", les avantages que l'utilisateur final devrait logiquement obtenir du fait de la modernisation du système.

A. MISE EN PLACE DE LA NOUVELLE VERSION DE CELEX

Le projet global de développement est décomposé en deux sous-projets, la partie textuelle et la partie analytique. Les travaux de construction des deux sous-projets seront partiellement menés de front. Ainsi, le prototype "partie analytique" sera développé et testé pendant la construction du sous-projet 1 dans le but de pouvoir commencer la construction réelle du sous-projet 2 dès la mise en production de la partie textuelle.

Le project plan peut donc se résumer ainsi :

1. Sous-projet 1 : Production des données textuelles

- Construction du texterface	4 h/m
- Construction de la BDPT ¹	1,5 h/m
- Mise en place de l'interface BDPT - BDD ²	1,5 h/m
- Test et mise en place de BDPT	5 h/m
	<hr/>
	12 h/m

2. Sous-projet 2 : Production des données analyse, diffusion et pilotage

- Construction de prototype	5,5 h/m
- Construction du module saisie/alimentation	4 h/m
- Construction de la BDPA ³	3,5 h/m

¹ Base de production des textes

² Base de diffusion

³ Base de production de la partie analytique

- Construction interface BDPA-BDD ²	6	h/m
- Construction de la BDD ²	4	h/m
- Test de mise en place de BDPA ³ et BDD ²	7	h/m
	<u>30</u>	h/m

Les ressources nécessaires au projet global se situeraient par conséquent près de 42 hommes/mois, tous investissements (SDA, SII, CC) confondus.

B. EVALUATION DE LA REDUCTION DES COÛTS D'UTILISATION

Sur base des coûts d'utilisation actuels d'une part et d'une évaluation des économies pouvant résulter d'une utilisation simplifiée du système issue de sa modernisation d'autre part, certaines hypothèses approximatives sont avancées ci-après avec les réserves d'usage. Une base d'évaluation plus scientifique de la réduction des coûts pourra être fournie dès achèvement des tests qui seront effectués sur un prototype.

1) Coûts de formation

Exemple : membre d'un cabinet juridique moyen situé à Frankfurt.

<u>CELEX</u>	<u>CELEX</u>	<u>BENE-</u>
<u>actuel</u>	<u>modernisé</u>	<u>FICE</u>

a) Formation à Bruxelles

- Cours de formation initial à Bruxelles	(2 j)	(1 j) ⁴
. droit d'inscription	4.000 FB	3.000 FB

⁴ pourra être réduit à 1 jour en raison de la simplification de l'accès.

. frais de voyage et hôtel	+ 15.000 FB	11.000 FB	+ 30 %
. manque à gagner	+ 16.000 FB	6.000 FB	
	<u>35.000 FB</u>	<u>20.000 FB</u>	

- Cours de perfectionnement	(1 j)	5	
. droit d'inscription	2.500 FB		
. frais de voyage et hôtel	11.000 FB		+ 100 %
. manque à gagner	8.000 FB		
	<u>21.000 FB</u>		

b) Formation sur place

(2 j)

. droit d'inscription (pour 10 élèves)	3.000 FB	6	+ 65 % ⁶
. manque à gagner	16.000 FB		
	<u>19.000 FB</u>		

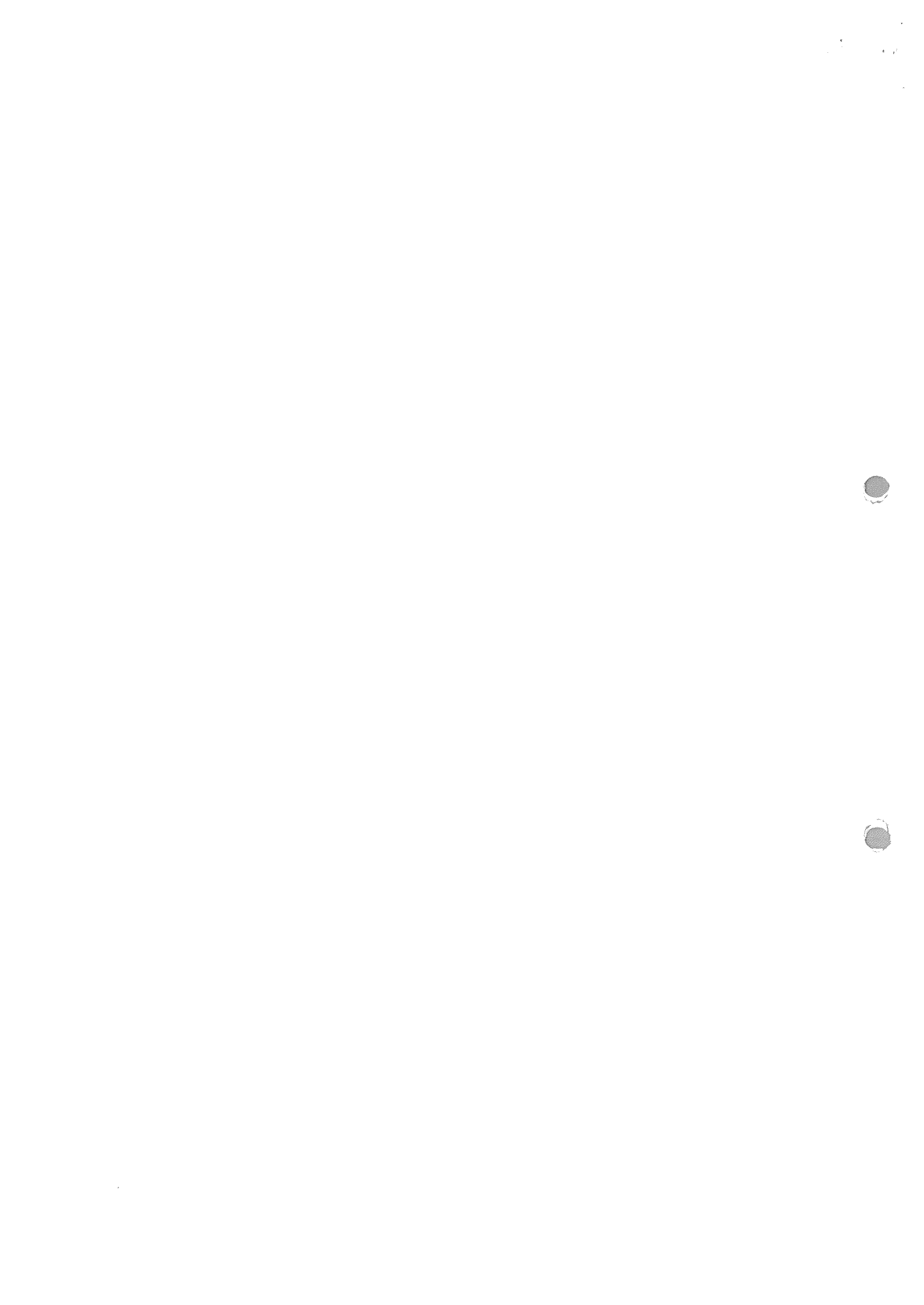
2. Coûts d'accès à l'information et l'exploitation des données

Les coûts de l'accès à l'information peuvent être chiffrés sur base du temps employé par l'utilisateur pour obtenir un résultat recherché en sachant que les prix de l'heure de connexion est de 4 200 FB et de l'heure de prestation de service de + 1 500 FB.

Dans l'hypothèse où un utilisateur occasionnel tel que décrit plus haut veut savoir quelles étaient par exemple, en 1983, les indications admises pour la désignation des vins de tables sur l'étiquetage (art. 3 du R/ 79/355), le scénario suivant peut être avancé :

⁵ pourra être supprimé en raison de la simplification de l'accès et de la disponibilité de cours sur support magnétique.

⁶ pourra être supprimé en raison de la disponibilité du cours sur support magnétique.



<u>PROCEDURE</u>	<u>TEMPS EMPLOYE</u>	<u>COUT DE CONNEXION</u>	<u>COUT DE PRESTATION DE SERVICE</u>
<u>CELEX actuel</u>			
- sélection et affichage de tous les documents "vins"	0,15 h	1 050 FB	375 FB
- sélection du sous-ensemble en vigueur en 1983	0,10 h	700 FB	250 FB
- sélection intellectuelle du seul texte pertinent par :			
. affichage et lecture des textes existants	0,50 h	3 500 FB	1 000 FB
. isolation des références de publication des textes manquants	0,15 h	1 050 FB	375 FB
. recherche des textes manquants dans JO et lecture	0,25 h	-	625 FB
- recherche des actes modificateurs des dispositions pertinentes de l'acte sélectionné	0,10 h	700 FB	250 FB
- reconstitution intellectuelle puis matérielle (copie, coupage, collage) des dispositions pertinentes et leur incorporation dans le document de travail	1,30 h	-	2 250 FB
TOTAL	3,25 h	7 000 FB	5 125 FB
COOT GENERAL		12 125 FB	

<u>PROCEDURE</u>	<u>TEMPS EMPLOYE</u>	<u>COUT DE CONNEXION</u>	<u>COUT DE PRESTATION DE SERVICE</u>
<u>CELEX modernisé</u>			
- sélection et affichage des seuls documents pertinents	0,20 h	1 400 FB	500 FB
- sélection du seul document en vigueur en 1983	0,10 h	700 FB	250 FB
- affichage de la version modifiée du texte et print des dispositions pertinentes	0,10 h	700 FB	250 FB
- incorporation dans le document de travail	0,10 h	-	250 FB
TOTAL	0,50 h	2 800 FB	750 FB
COOT GENERAL		3 550 FB	

Ces chiffres montrent clairement que les bénéfices que les utilisateurs pourraient retirer de la modernisation du système sont appréciables. Il faut cependant souligner à nouveau que l'évaluation du temps de recherche dans le système modernisé est théorique. Elle se base sur les déductions suivantes qui sont à confirmer par les tests du prototype :

- la durée de recherche devrait être réduite par la simplification de la structure et par la suppression, grâce au guidage, du tâtonnement de l'utilisateur (résultats partiels, stratégies de recherche inadéquates, etc.);
- l'existence d'une version codifiée et mise à jour des actes supprimerait l'examen long et fastidieux des nombreux actes modificateurs postérieurs;
- la couverture exhaustive en texte intégral éviterait un recours complémentaire au Journal officiel et la meilleure présentation des textes (typographie riche, formatage clair, etc.) permettrait d'incorporer les données directement dans un document de travail ou un dossier, sans qu'il soit nécessaire de re-dactylographier les données.

NOUVEAUX CHAMPS ET LEXIQUES CELEX

NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB	NUMERO	ND	Numéro d'identification standardisé du document (SAATNNNN(nn))	TOUS SECTEURS ¹ Alphanumérique, vocabulaire contrôlé
	TYPE	DT	Nature juridique du document symbolisée par une lettre (R = Règlement, D = Décision)	TOUS SECTEURS Alphanumérique, vocabulaire contrôlé Synonymie entre la lettre du numéro et le libellé long de la forme juridique de l'acte
	AUTEUR	AU	Nom de l'institution, organe, pays dont émane l'acte; nome de la personne physique (parlementaire, Avocat général, etc.) auteur du document	TOUS SECTEURS Alphanumérique, vocabulaire contrôlé Synonymie et relation de groupe
	REF.PUB	S0	Références de publication au J.O. ou au RECUEIL (AAAA/MM/JJ)	TOUS SECTEURS Alphanumérique, texte

NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	DESCRIPT	CT	Descripteurs du contenu du document	TOUS SECTEURS Alphanumérique, vocabulaire contrôlé Synonymie et relation de groupe Thésaurus standardisé (peut-être EUROVOC ?)
	REQUER	-	Nom de la partie requérante dans une affaire devant la C.J. (institution, organe, pays, personne physique ou morale)	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	DEFEND	-	Nom de la partie défenderesse	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	INTERV	-	Nom de la partie intervenante	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	OBSERV	-	Institution, organe ou état membre ayant formulé des observations dans une affaire devant la C.J.	SECTEUR 6 Alphanumérique, vocabulaire contrôlé

NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	NATIONA	-	Nationalité du requérant ou défendant particuliers ou du membre du PE Pays d'appartenance du tribunal dont émane la mande de décision pré-judicielle	SECTEURS 6 et 9 Alphanumérique, vocabulaire contrôlé Synonymie
	TRIB.NAT	-	Références des jugements des tribunaux nationaux ayant donné lieu à un renvoi pré-judicielle	SECTEUR 6 Alphanumérique, texte
	AVOC.G	-	Nom de l'avocat général	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	JUGE.RAP	-	Nom du juge rapporteur	SECTEUR 6 Alphanumérique, vocabulaire contrôlé
	PROCEDUR	-	Nature du recours, par exemple : (recours de fonctionnaire; recours en annulation, en carence, en constatation de manquement, demande en intervention, etc.)	SECTEUR 6 Alphanumérique, vocabulaire contrôlé Synonymie

NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	REPERT	-	Code numérique de classification dans le "Répertoire de la législation communautaire en vigueur"	SECTEUR 1 à 7 Alphanumérique, vocabulaire contrôlé Synonymie entre le code numérique et les descripteurs correspondant à la classification par matières de l'activité législative Relation de groupe
	LANGUE	LA	Langue(s) faisant foi; langue de procédure	SECTEURS 1 à 4, 6 à 8, 10 Alphanumérique, vocabulaire contrôlé Synonymie
	DESTINA	-	Etats, institutions ou particuliers auxquels l'acte est adressé	SECTEURS 2 à 5 Alphanumérique, vocabulaire contrôlé Synonymie
	DEPOS	-	Organe dépositaire des accords	SECTEUR 2 Alphanumérique, vocabulaire contrôlé Synonymie
	DISP XX	-	12 champs contenant les dispositions nationales d'exécution des directives communautaires (XX = BE, DA, DE, EL, ES, FR, IR, IT, LU, NL, PO, UK) avec références de publication dans les JO nationaux	SECTEUR 3L Alphanumérique, texte

NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
BIB (suite)	INFO.DIV	-	Informations diverses, références internes des documents, commentaires utiles	SECTEURS 2 à 7 Alphanumérique, vocabulaire contrôlé
REL	BASE	-	Numéro d'identification des documents constituant la base juridique de l'acte examiné	SECTEUR 2 à 7 Alphanumérique, texte Synonymie
	CIT	-	Numéro d'identification des documents cités dans l'acte concerné	SECTEURS 2 à 7 Alphanumérique, texte
	MOD	-	Actes cités, visés, modifiés ou affectés par l'acte concerné	SECTEUR 1 à 7 et 9 Alphanumérique, texte avec indexation positionnelle
	ULT	-	Actes qui visent, modifient ou affectent l'acte concerné	SECTEUR 1 à 7 et 9 Alphanumérique, texte
	TRA.PREP	-	Références des actes préparatoires (propositions, avis) qui sont à l'origine de l'acte concerné	SECTEUR 1 à 7 Alphanumérique, texte
	NOTES	-	Références aux articles parus dans des revues juridiques relativement à l'arrêt concerné	SECTEUR 6 Alphanumérique, texte

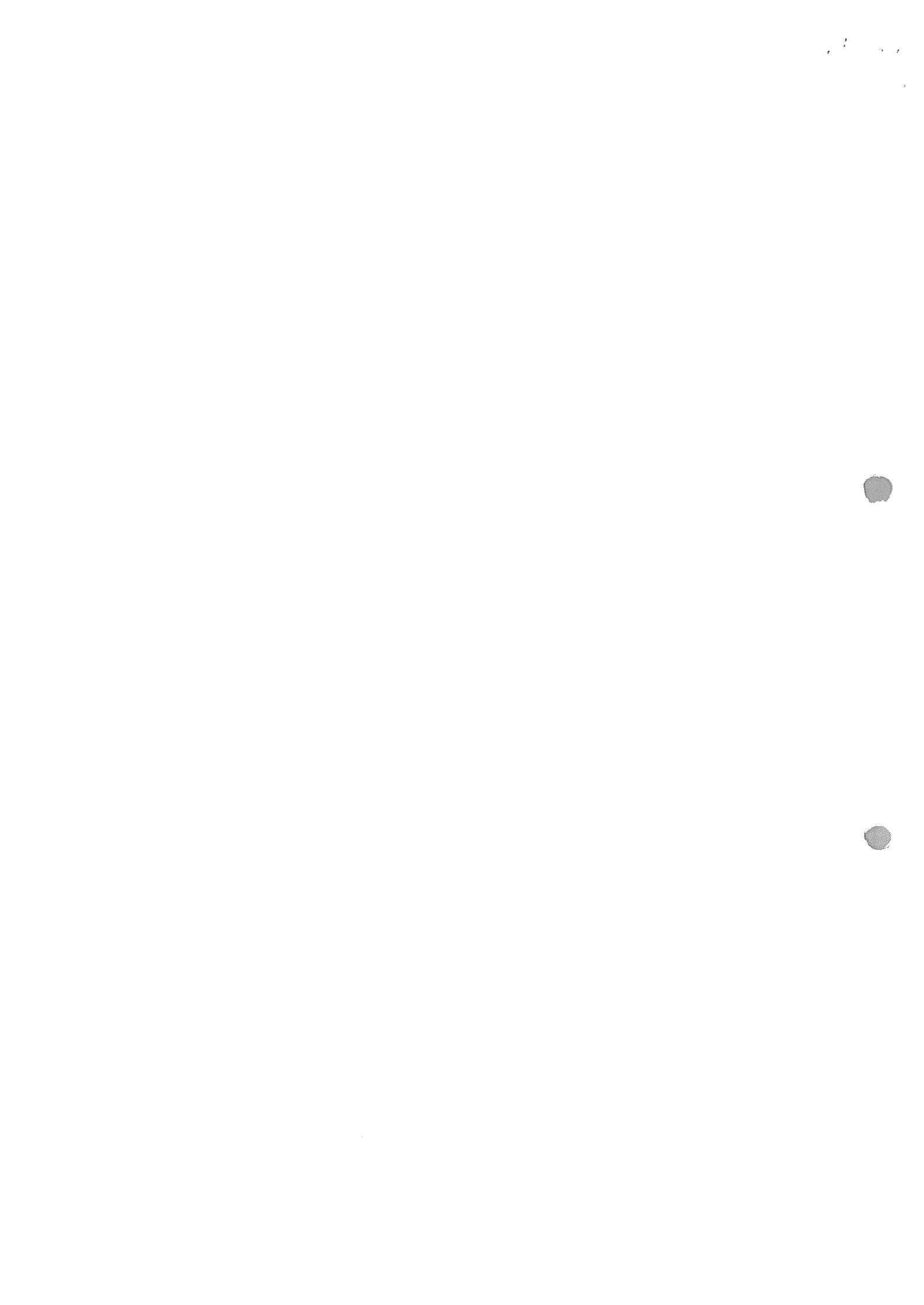
NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
DATE	DOC	-	Date d'adoption de l'acte, date du prononcé de la décision de la Cour (YYYY/MM/DD)	TOUS SECTEURS Date
	DEM	-	Date de dépôt d'une affaire au Greffe de la Cour (YYYY/MM/DD)	SECTEURS 6 et 8 Date
	EFFET	-	Date d'entrée en vigueur (YYYY/MM/DD ENTREE EN VIGUEUR) Date de mise en application (YYYY/MM/DD IMPLEMENTATION) Date de notification (YYYY/MM/DD NOTIFICATION)	SECTEURS 1 à 5 Date, avec caractéristique
	FIN.VAL	-	Date de fin de validité	SECTEUR 1 à 5 Date
	TRANSPOS	-	Date à laquelle les Etats membres doivent se conformer à une directive	SECTEUR 3 Date
	SESSION	-	Date du débat d'une résolution ou d'un avis du PE (YYYY/MM/DD DEBAT) Date de vote d'une résolution ou d'un avis (YYYY/MM/DD VOTE)	SECTEURS 5 et 9 Date, avec caractéristique

NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
TXT	TITRE	TI	Titre officiel du document (dans le JO, dans le Receuil, dans les journaux officiels des EEMM, etc.)	TOUS SECTEURS Alphanumérique, texte avec indexation positionnelle
	VISA	-	Dispositions visées dans l'acte concerné	SECTEURS 1 à 7 Alphanumérique, texte avec indexation positionnelle
	CONSIDER	-	Considérations en fait ou en droit précédant le texte du document	SECTEURS 1 à 5 et 7 Alphanumérique, texte avec indexation positionnelle
	CORPUS	-	Partie textuelle proprement dite du document	SECTEURS 1 à 5, 7 et 9 Alphanumérique, texte avec indexation positionnelle
	FOOTNOTE	-	Note en bas de page du document	SECTEURS 1 à 5 et 7 Alphanumérique, texte avec indexation positionnelle
	ANNEXE	-	Annexes du document	SECTEURS 1 à 5 et 7 Alphanumérique, texte avec indexation positionnelle

NOM DU LEXIQUE	NOM DU CHAMP	NOM DU CHAMP EN CCL	CONTENU DU CHAMP	COUVERTURE ET DESCRIPTION DU CHAMP
TXT (suite)	RECTIF	-	Texte des rectificatifs de l'acte concerné	SECTEURS 1 à 5 Alphanumérique, texte avec indexation positionnelle
	SOMMAIRE	-	Résumé de l'arrêt tel que publié au Recueil de la jurisprudence	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	INDEX	-	Mots clés attribués à l'arrêt, tels que publiés Recueil	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	INITITULE	-	Nom des parties, objet du recours, composition de la Cour	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	MOTIFS	-	Texte de la partie en droit de l'arrêt	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	DEPENS	-	Texte de la décision de la Cour sur les dépenses	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	DISPOSI	-	Texte du dispositif de l'arrêt	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle
	CONCLUS	-	Texte des conclusions de l'avocat général	SECTEUR 6 Alphanumérique, texte avec indexation positionnelle

SECTEURS DU NOUVEAU CELEX

1. TRAITES
2. RELATIONS EXTERIEURES
3. DROIT DERIVE ET DISPOSITIONS NATIONALES D'EXECUTION DES DIRECTIVES
4. DROIT COMPLEMENTAIRE
5. TRAVAUX PREPARATOIRES
6. JURISPRUDENCE DE LA COUR DE JUSTICE DE LA C.E.
7. VERSIONS DES TEXTES
8. JURISPRUDENCE NATIONALE RELATIVE AU DROIT COMMUNAUTAIRE
9. QUESTIONS PARLEMENTAIRES
10. DOCTRINE RELATIVE AU DROIT COMMUNAUTAIRE



LES GRANDS TRAITS DU COMMUN COMMAND LANGUAGE (CCL)¹

Le CCL est un langage de commande standardisé qui facilite les recherches en ligne en mode conversationnel par l'emploi de noms, de commande significatifs et de leur abréviations d'une part, et d'autre part, par une syntaxe et une sémantique de langage simples dans leur utilisation.

1. Principales commandes

Le choix de la base de données s'opère par la commande BASE ou B;

Les données se visualisent par la commande SHOW ou S et s'impriment avec la commande PRINT ou P;

La commande DEFINE ou DEF permet de spécifier une édition ou un lexique par défaut;

L'historique de la recherche est affiché par la commande REVIEW ou R;

DELET ou DEL supprime une ou plusieurs étapes de recherche;

Les commandes UST et DISPLAY affichent l'environnement alphabétique et relationnel d'un terme;

BACK et MORE font remonter ou descendre l'écran;

Les commandes ? ou HELP assistent l'utilisateur et STOP arrête la session d'interrogation;

La commande FIND ou F, enfin, permet de rechercher directement les données pertinentes; elle est suivie par l'introduction d'éléments et d'opérateurs de recherche.

2. Syntaxe et opérateurs

Les opérateurs booléens sont AND, OR, NOT; les opérateurs de relations sont : = /. La virgule sépare les termes de recherche et les signes +, !, ?, = (opérateurs de distance) masquent un nombre différent de caractères.

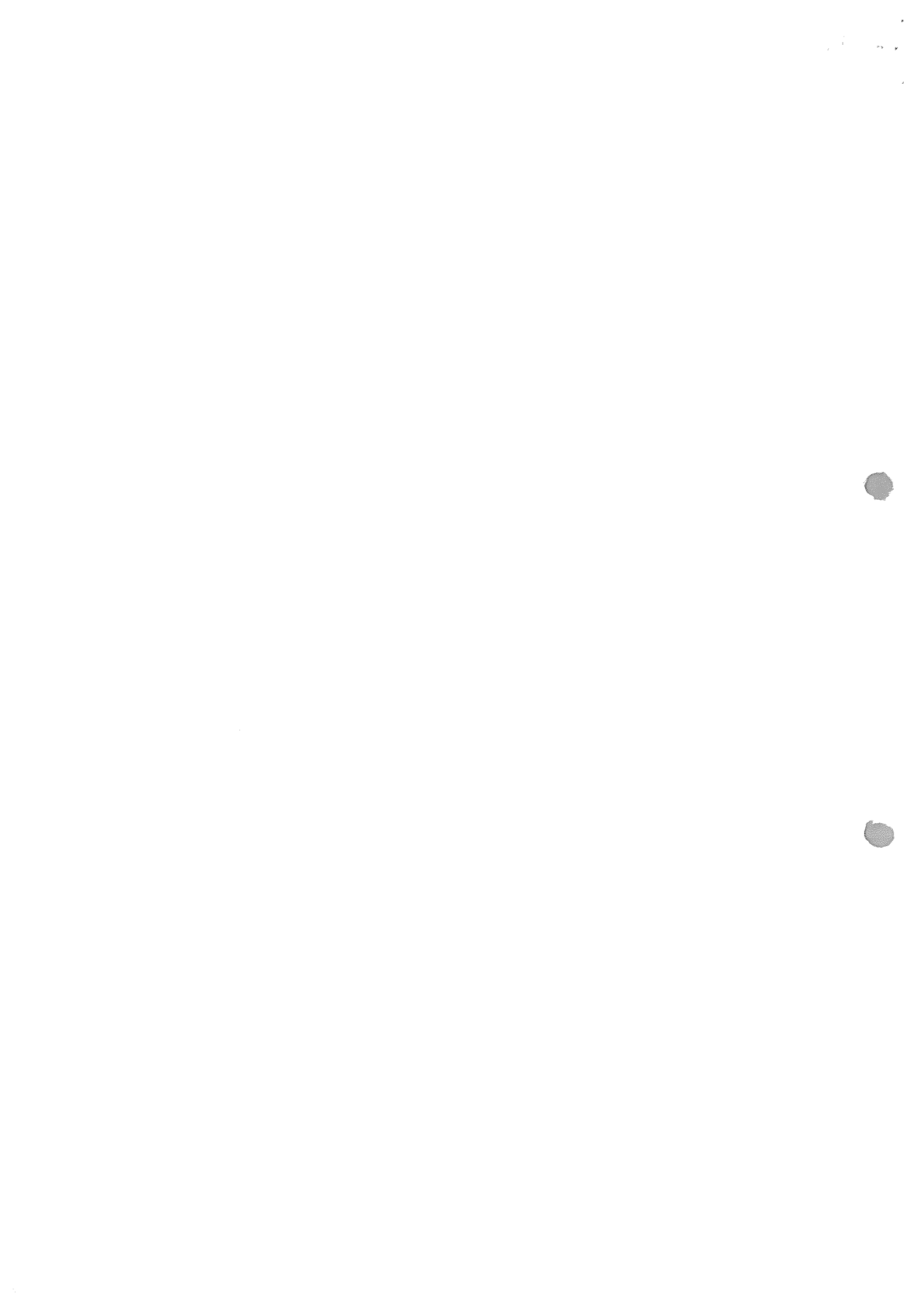
La syntaxe permet de combiner,² dans une même étape de recherche des recherches sur plusieurs champs ainsi que différents opérateurs, sans ordre de priorité.

Pour limiter la recherche à un champ, il faut faire suivre le terme de recherche par un trait oblique et par le nom du champ.

Ex. : F CONSEIL/AUTEUR AND R/TYP.DOC

¹ Version ISO/DIS 8777

² Les noms des champs de base standardisés sont mnémoniques en anglais. Ex. : AU = auteur, DT = type de document etc.



Le terme ALL précédant le terme de recherche permet de travailler en mode transactionnel pour les termes multiples.

Ex. : F ALL 03+/REPERT

On trouvera une description in extenso des divers détails du langage commun de commande que le présent aperçu n'avait pas pour but de développer le dominant DRAFT INTERNATIONAL STANDARD ISO/DIS 8777.

