

E.N.S.S.I.B.
ECOLE NATIONALE SUPERIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES

UNIVERSITE
CLAUDE BERNARD
LYON 1

DESS en INFORMATIQUE DOCUMENTAIRE

Rapport de stage

**Intégration du langage SGML dans l'édition scientifique
et technique au Cemagref**

Valérie Bédel

Sous la direction de
Jacques Clément Chef du Service IST
Camille Cédra Editeur

Cemagref

Direction de la Communication et de la Valorisation
BP 22, 92162 ANTONY Cedex - Tél (1) 40 96 61 21 - Télécopie (1) 40 96 61 39

1996

Titre : L'intégration du langage SGML dans l'édition scientifique et technique

Auteur : Valérie Bédel

Résumé :

A travers un exemple concret : les actes de colloque publiés par le Cemagref, ce rapport décrit la faisabilité de l'intégration du langage SGML dans l'édition scientifique et technique. Successivement, sont abordées : la définition de du langage SGML et l'écriture d'une DTD (Document Type Definition) pour les actes de colloque. Pour chaque maillon de la chaîne éditoriale, des solutions d'intégration sont proposées.

Mots-clés :

SGML - Edition électronique - Edition scientifique et technique - balisage de document - structure logique - Acte de colloque

Abstract :

The feasibility of the application of SGML language to scientific and technical publications is described in this report taking as a basis a concrete example i.e proceedings published by Cemagref. Defining SGML language, writing a DTD (Document Type Definition) for Cemagref proceedings are the main topics dealt with. Various SGML applications are put forward for each publishing step.

Keywords :

SGML - Electronic publishing - Scientific and technical publishing - Tag - Logical structure - Proceedings

Remerciements

Je tiens à remercier le Service IST de m'avoir offert la possibilité de poursuivre une mission si enrichissante, qui m'a permis de compléter et d'approfondir ma formation et m'a offert la possibilité d'appliquer mes connaissances sur le terrain.

Je remercie chaleureusement Jacques Clément et Camille Cédra de l'encadrement sérieux et régulier qu'ils m'ont accordé, ainsi que pour leurs précieux conseils et leur soutien. J'ai grandement apprécié de bénéficier d'une telle autonomie quant à la gestion de mon travail. Je les remercie également d'avoir mis à ma disposition tous les outils nécessaires au bon déroulement de ma mission, et de la confiance qu'ils ont bien voulu m'accorder.

Je remercie Julienne Baudel et Karim Haddane d'avoir participé au suivi de mon stage.

Enfin je tiens à remercier Delphine Challeil, Maurice Merlin, Françoise Cédra, Martine Boudot-Lamotte et Nicole Delherbe pour leurs conseils et leur accueil chaleureux au sein du service IST (Information Scientifique et Technique).

INTRODUCTION	4
PREMIERE PARTIE : <i>LE CEMAGREF</i>	5
1. Présentation du Cemagref.....	6
1.1 Le statut.....	6
1.2 Les moyens.....	6
1.3 Les missions.....	6
1.3.1 La recherche.....	6
1.3.2 Le transfert des connaissances.....	6
1.3.3 La diffusion.....	7
1.4 L'organisation du Cemagref.....	7
1.5 Localisation.....	8
2. Le service de l'Information Scientifique et Technique (IST)	9
2.1 L'organisation du service.....	9
2.1.1 La gestion et la diffusion de l'information.....	9
2.1.2 Publications, colloques et salons professionnels.....	10
2.1.3 L'édition.....	10
2.2 Les partenaires externes.....	10
2.3 Les moyens.....	10
2.3.1 Les moyens humains.....	10
2.3.2 Les moyens financiers.....	11
DEUXIEME PARTIE : <i>LE PROJET PROPOSE</i>	12
1. Présentation de la mission	14
1.1 L'intitulé.....	14
1.2 Le contexte.....	14
1.3 Les objectifs.....	16
2. Le langage S G M L	17
2.1 Historique et principes.....	17
2.2 Les avantages liés à l'utilisation de SGML.....	17
2.3 Le langage SGML.....	18
2.3.1 La notation SGML.....	18
2.4 Les composantes de SGML.....	18
2.4.1 La déclaration SGML.....	18
2.4.2 La DTD (Document Type Definition).....	18
2.4.3 L'instance de document.....	19
2.5 Les Outils SGML.....	19

2.5.1 Les éditeurs SGML.....	19
2.5.2 Les parseurs.....	20
2.6 L'installation SGML.....	20
TROISIEME PARTIE : LA METHODOLOGIE UTILISEE	21
1. Analyse préalable pour la conception de la DTD	23
1.1 Définition de la chaîne éditoriale.....	23
1.1.1 Les principes.....	23
1.1.2 Les acteurs et leur(s) rôle(s).....	23
1.2 Liste des applications envisagées.....	24
1.3 Les utilisateurs.....	24
2. Conception de la DTD	25
2.1 Constitution d'un corpus d'actes de colloque.....	25
2.2 Analyse du corpus et repérage de la structure logique.....	25
2.3 Les principaux éléments de la DTD congrès.....	28
2.4 Comparaison de la DTD congrès avec la DTD Griseli et la notice documentaire du Cemagref.....	32
2.4.1 Comparaison avec la notice du Cemagref.....	32
2.4.2 Comparaison avec la DTD Griseli.....	33
3. Ecriture de la DTD	34
3.1 Règles d'écritures.....	34
3.2 Ecriture et vérification.....	35
QUATRIEME PARTIE : INTEGRATION DE LA DTD DANS LA CHAINE EDITORIALE DU CEMAGREF..	37
1. SGML et la chaîne éditoriale	39
2. La production de texte	41
2.1 Les éditeurs SGML.....	41
2.2 Les différentes solutions d'intégration de SGML.....	43
2.3 Quel choix ?.....	44
2. L'impression	45
3. La diffusion en ligne	46
4. l'extraction des notices documentaires	47
5. SGML intégré dans la chaîne éditoriale	48

6. Eléments de réponse aux questions posées dans le cadre de la mission....	49
CONCLUSION	51
BIBLIOGRAPHIE.....	52
LISTE DE RESSOURCES SGML SUR INTERNET.....	55
ANNEXES	57

Introduction

L'objet du stage qui m'a été proposé est d'étudier la possibilité d'intégration du langage SGML (Standard Generalized Markup Language) pour les éditions du Cemagref. Cette étude s'est appuyée sur une application concrète : les actes de colloque.

Ce stage s'est déroulé au Cemagref d'Antony, au sein du service IST (Information Scientifique et Technique).

Après une présentation rapide du Cemagref et du service IST, nous allons exposer en détail le projet proposé et faire un rappel sur ce qu'est SGML. Dans la troisième partie, nous expliquerons la méthodologie utilisée pour écrire la DTD (Document Type Definition) des actes de colloque. Enfin, la dernière partie sera consacrée aux différentes solutions retenues pour l'intégration du langage SGML au sein des éditions du Cemagref.

PREMIERE PARTIE

Le Cemagref

1. Présentation du Cemagref

1.1 LE STATUT

Créé en 1981, le Cemagref est un Etablissement Public à caractère Scientifique et Technologique (EPST) placé sous la double tutelle des Ministères chargés de la Recherche et de l'Agriculture.

1.2 LES MOYENS

Le Cemagref emploie en permanence 1 000 personnes dont 480 ingénieurs et chercheurs auxquels s'ajoutent 100 thésards et 500 stagiaires de longue durée.

En 1996 son budget consolidé était de 325 millions de francs.

1.3 LES MISSIONS

1.3.1 La recherche

Le Cemagref mène des recherches finalisées dans trois domaines : les eaux continentales, les équipements agricoles et alimentaires et, la gestion des territoires.

Pour mener à bien ses recherches, le Cemagref est organisé en quatre départements scientifiques :

- ⇒ Gestion des milieux aquatiques,
- ⇒ Gestion des territoires,
- ⇒ Equipements pour l'eau et l'environnement,
- ⇒ Equipements agricoles et alimentaires.

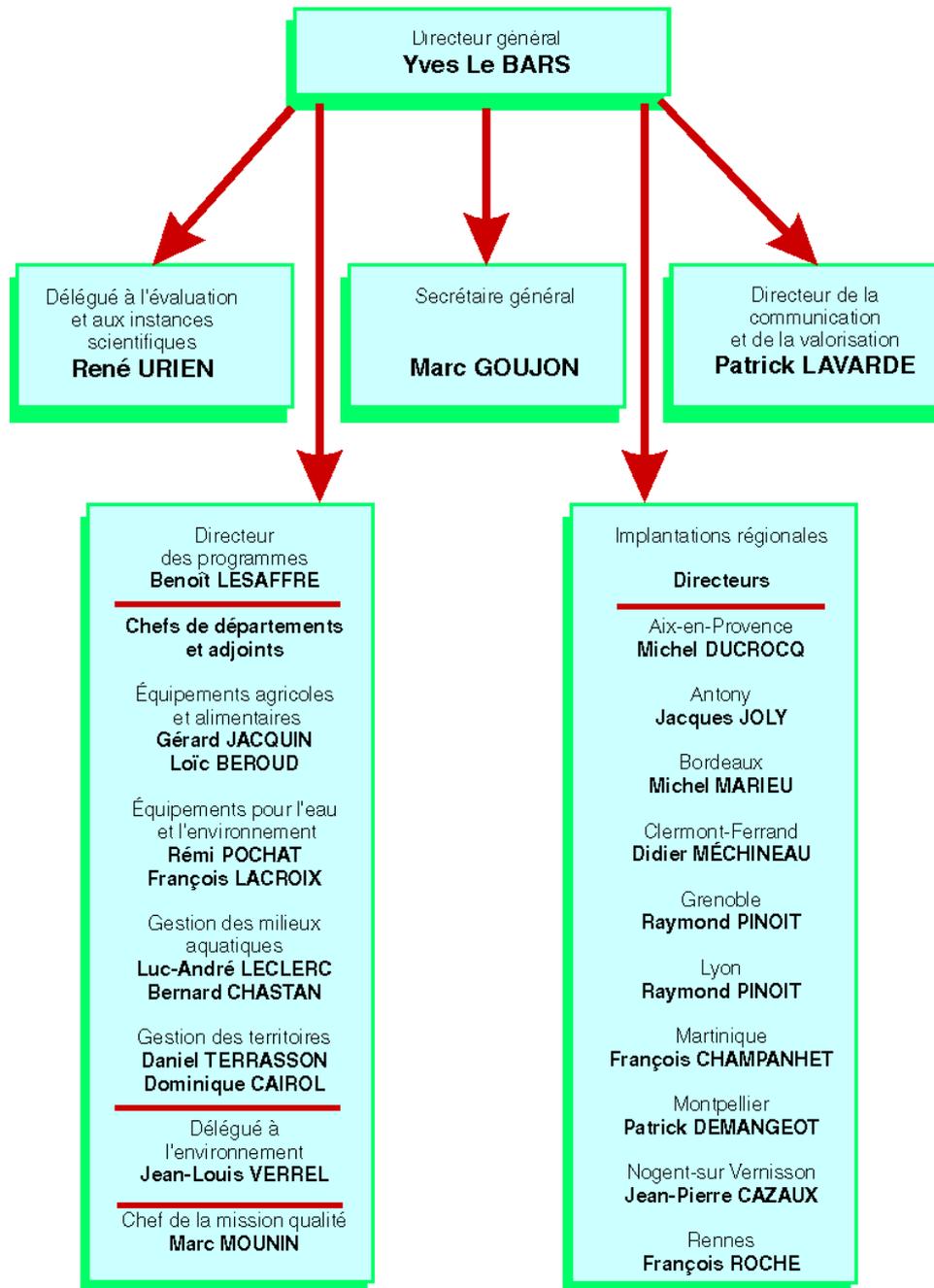
1.3.2 Le transfert des connaissances

Le Cemagref valorise ses résultats auprès de la communauté scientifique, des industriels, des bureaux d'études, et dépose des brevets. Il est également très impliqué auprès des universités et des grandes écoles.

1.3.3 La diffusion

Pour valoriser les résultats de ses travaux de recherche, le Cemagref édite des publications scientifiques et techniques étroitement liées aux champs d'application de la recherche de l'établissement.

1.4 L'ORGANISATION DU Cemagref



Le directeur général assure la direction scientifique, technique, administrative et financière de l'établissement.

Il est assisté par :

- ⇒ Une direction scientifique qui définit et met en oeuvre le fonctionnement scientifique de l'établissement et assure l'animation scientifique. Elle est composée d'un délégué aux évaluations et aux instances scientifiques, d'un directeur des programmes et des chefs de départements,
- ⇒ Un secrétaire général chargé de la gestion administrative et financière,
- ⇒ Un directeur de la communication et de la valorisation qui prépare et met en oeuvre la politique de communication et d'information scientifique et technique,
- ⇒ Les directeurs de groupements.

1.5 LOCALISATION

Le Cemagref est composé de dix groupements répartis en dix sites géographiques : Aix-en-Provence, Antony, Bordeaux, Clermont-Ferrand, Grenoble, Lyon, La Martinique, Montpellier, Nogent-sur-Vernisson et Rennes.

La Direction Générale du Cemagref est située à Antony.

2. Le service de l'Information Scientifique et Technique (IST)

En tant qu'établissement de recherche, le Cemagref est producteur mais aussi utilisateur d'information scientifique et technique.

La loi d'orientation de la recherche de 1982 assigne aux établissements publics scientifiques et techniques cinq missions :

- ⇒ la production,
- ⇒ le transfert des connaissances,
- ⇒ la participation à la formation initiale et continue,
- ⇒ l'administration de la recherche,
- ⇒ la diffusion de l'information scientifique et technique.

C'est pour remplir cette dernière mission que le Cemagref a mis en place un service de l'Information Scientifique et Technique. La politique de l'IST a pour objectif de mieux faire connaître le Cemagref et ses activités de recherche. A ce titre, l'établissement distingue trois types d'information : scientifique, technique et la vulgarisation scientifique. Ces informations sont destinées à trois publics cibles : les scientifiques, le milieu socioprofessionnel, le grand public. Les deux premières cibles sont prioritaires, la dernière sert surtout à valoriser l'image du Cemagref.

2.1 L'ORGANISATION DU SERVICE

Le service IST dépend de la DICOVA (Direction de la Communication et de la Valorisation) qui appartient à la Direction Générale. La DICOVA regroupe également les Services des Relations Européennes et Internationales, des Relations avec les Entreprises, et le Service de la Communication.

Créé au premier janvier 1996, le service IST est né de la fusion des services "Valorisation Scientifique et Technique" et "Editions". Il est composé de trois entités :

- ⇒ La gestion et la diffusion de l'information,
- ⇒ Les publications, colloques et salons professionnels,
- ⇒ L'édition.

2.1.1 La gestion et la diffusion de l'information

Cette entité s'occupe de la documentation : animation du réseau des documentalistes répartis dans les différents groupements géographiques, gestion de la documentation de la Direction

Générale, enrichissement de la photothèque et développement et mise à jour des services d'information en ligne (serveur Web, Minitel).

2.1.2 Publications, colloques et salons professionnels

Il s'agit ici de l'organisation de sessions de formation à l'écriture scientifique, de la participation de chercheurs du Cemagref à des revues scientifiques, du dépôt et de la valorisation des publications du Cemagref ainsi qu'une assistance à l'organisation de la participation à des colloques.

2.1.3 L'édition

L'existence de l'entité édition se justifie par le fait que le Cemagref ne peut s'appuyer entièrement sur les éditeurs privés en raison de la spécificité des sujets abordés dans les ouvrages et de leur faible tirage. Cette entité fonctionne comme un service d'édition privée, cependant la priorité est donnée à l'édition d'ouvrages rédigés par les auteurs du Cemagref. La programmation éditoriale fait partie intégrante de la programmation scientifique. En moyenne, vingt livres sont édités par an.

Il est à noter que l'ensemble des supports éditoriaux est utilisé. En 1995, le Cemagref a édité deux CD-Rom. L'édition papier reste néanmoins majoritaire.

En plus des ouvrages, le Cemagref édite une revue trimestrielle : *INGENIERIES-EAT*.

2.2 LES PARTENAIRES EXTERNES

Le service IST travaille en collaboration avec des structures IST, des organismes de recherche et, des producteurs de bases de données tel que l'INIST (Institut de l'Information Scientifique et Technique).

2.3 LES MOYENS

2.3.1 Les moyens humains

Le service IST emploie 16 personnes qui sont réparties de la manière suivante :

Gestion et diffusion de l'information	5
Publications, colloques, salons professionnels	2,5

Editions	6
Commun au service	1,5
Commun à l'établissement	1
TOTAL	16

2.3.2 Les moyens financiers

Le budget de 1997 est de 5 millions de francs composés de 1,4 millions de ressources propres et 3,3 millions de francs de contribution de l'établissement. Les dépenses sont réparties de la manière suivante :

Gestion et diffusion de l'information	1,3 MF
Publications, colloques, salons professionnels	1,4 MF
Editions	2 MF

DEUXIEME PARTIE

Le projet proposé

Dans cette partie, nous allons expliquer en détail l'objet de la mission : son intitulé, ses objectifs définis par le service IST et le contexte dans laquelle elle s'inscrit.

Ensuite, nous rappellerons la définition du langage SGML, ses composantes ainsi que les différents types d'outils qui sont associés à ce langage.

1. Présentation de la mission

1.1 L'INTITULE

Le projet proposé par le service IST avait pour intitulé :

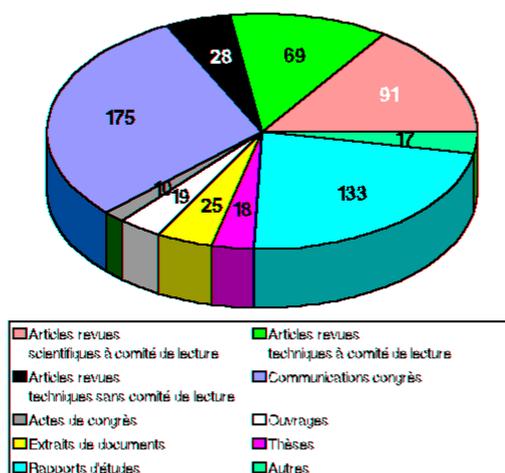
"L'intégration du langage SGML dans la gestion de l'édition scientifique et technique d'un établissement public de recherche :

- ⇒ typologie des ouvrages du Cemagref, analyse et évaluation de l'opportunité de leur structuration selon la norme SGML,
- ⇒ méthodologie et itinéraires d'intégration du balisage SGML pour les documents existants ou à éditer par la suite...
- ⇒ application concrète conduisant à la proposition d'une DTD pour une série homogène de documents."

1.2 LE CONTEXTE

◆ Les publications

En 1995, les scientifiques du Cemagref ont rédigé 585 publications qui sont réparties de la manière suivante :



Au travers de ce schéma, on se rend compte que les communications à un congrès sont le type de publication le plus rédigé par les ingénieurs-chercheurs du Cemagref. Le deuxième type de publication est les rapports d'études. Le Cemagref a donc une production de littérature grise importante.

Par rapport aux missions attribuées aux établissements publics de recherche, un des soucis du Cemagref est de valoriser au mieux l'ensemble de ses publications. La valorisation des documents passe notamment par un meilleur recensement dans les bases de données. Pour améliorer ce recensement, le Cemagref a décidé de participer au projet Griseli dont la mise en oeuvre incombe principalement à l'INIST.

Griseli est un programme de collecte, traitement et communication de la littérature grise française mis en oeuvre par le Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche et la DISTB (Direction de l'Information Scientifique et Technique et des Bibliothèques). L'INIST est le guichet national pour ce projet. Dix pôles y participent, chacun représentant un domaine scientifique. Ils sont chargés de fournir à l'INIST les notices SGML des documents appartenant à la littérature grise du domaine concerné. Ces notices doivent respecter la DTD (Document Type Definition) fournie par l'INIST.

Le Cemagref est donc chargé de fournir à l'INIST d'une part les notices documentaires des publications des scientifiques de l'établissement et d'autre part les notices des ouvrages édités par le Cemagref.

◆ L'édition :

Au sein de ses 585 publications, le Cemagref édite en moyenne 20 ouvrages par an parmi lesquels 5 actes de colloque. Les actes de colloque représentent donc 20% de la production éditoriale.

En moyenne, ils sont édités à 500 exemplaires.

◆ La documentation

La gestion documentaire du Cemagref est décentralisée. Chaque groupement dispose de son propre service de documentation et de sa propre base de données. Chaque site a la possibilité d'accéder à la base de données locale et aux bases des autres groupements. En effet, les services de documentation disposent du même logiciel documentaire : Texto pour Windows et, utilisent un format de notice commun. De plus, le vocabulaire utilisé pour l'indexation est issu du thesaurus du Cemagref.

En plus de ces bases de données, il existe la base Cemagref-Publications qui recense les publications rédigées ou co-rédigées par les ingénieurs-chercheurs du Cemagref. A cet effet, la DICOVA a mis en place un dépôt centralisé des publications.

1.3 LES OBJECTIFS

Les objectifs de la mission par ordre de priorité sont :

- ⇒ Démontrer la maniabilité d'une application SGML dans le cadre de l'édition scientifique et technique, étudier les problèmes de mises en oeuvre (y compris les problèmes de coût au niveau matériel et humain) que cela implique et identifier les contraintes.
- ⇒ Démontrer si cela améliore la productivité pour l'édition des actes de colloque.
- ⇒ Améliorer la diffusion de la littérature grise du Cemagref en multipliant les supports de diffusion (papier, CD-Rom, en ligne). La littérature grise, au Cemagref, représente 60% de la production de textes. Cette amélioration passe notamment par une diffusion en ligne.
- ⇒ A travers une norme indépendante des moyens matériels, assurer un rôle de mémoire et de pérennité de l'information.
- ⇒ Etudier l'acceptabilité de SGML pour les auteurs des communications. Ceux-ci devront-ils avoir des compétences spécifiques ?

Les actes de colloque ont donc été choisis car ils constituent une part importante à la fois des publications et des éditions du Cemagref mais aussi parce qu'ils présentent une forte structuration avec une grande cohérence d'une communication à l'autre. De plus, leur édition doit être assez rapide pour qu'ils puissent être disponibles le jour du colloque. Enfin, au travers des actes de colloque deux contraintes apparaissent : la structuration des communications et la structuration de l'ouvrage en lui-même.

Pour construire la DTD (Déclaration de Type de Document), on se basera sur deux DTD existantes :

- ⇒ la DTD de l'INIST appartenant au projet Griseli, la DTD créée devra être compatible avec la DTD du projet Griseli, ceci afin de pouvoir fournir automatiquement les notices SGML des actes de colloque et des communications à l'INIST.
- ⇒ La DTD normalisée des documents scientifiques ISO 12083, afin d'avoir un système ouvert et normalisé.

La conception de la DTD donnera l'occasion de faire une analyse du marché des outils SGML existants.

2. Le langage SGML

2.1 HISTORIQUE ET PRINCIPES

La norme SGML ou Langage Normalisé de Balisage Généralisé a été créée par un groupe de recherche de l'ISO en 1986 afin de faciliter les échanges de documents électroniques.

SGML est un métalangage c'est à dire un moyen de description d'un langage. Il ne s'intéresse pas directement à l'aspect final des documents (mise en page, typographie...) mais à leurs formes retraitables. Ainsi un document qui est balisé en SGML pourra être édité sur différents types de supports. SGML permet donc de décrire la structure logique d'un document et plus précisément d'une classe de documents comme les livres ou encore les publications en série.

Le but de SGML est de pouvoir créer des documents échangeables et révisables. En effet, cette norme a été conçue pour être totalement indépendante des plate-formes matérielles utilisées.

Pour l'instant, SGML est essentiellement utilisé dans les milieux de l'édition et plus particulièrement pour l'édition de la documentation technique qui contient une information très structurée et qui nécessite de nombreuses mises à jour. Ainsi, le Ministère de la Défense américain (U.S Department of Defense : Dod) utilise SGML dans le cadre du projet CALS (Continuous Acquisition and Life Support) afin de pouvoir stocker, échanger et distribuer de manière électronique toute sa documentation technique en provenance et à destination de ses différents fournisseurs.

2.2 LES AVANTAGES LIES A L'UTILISATION DE SGML

⇨ Longévité ou pérennité de l'information :

Etant indépendante des logiciels et moyens matériels utilisés, les documents balisés en SGML peuvent être toujours utilisables.

⇨ Augmentation de la productivité :

SGML permet de décrire la structure logique des documents sans tenir compte de la structure physique, ceci permet donc d'éditer un même document avec différentes mises en forme.

⇨ Portabilité améliorée de l'information :

L'indépendance vis à vis des plates-formes matérielles permet l'échange de documents d'un ordinateur à l'autre sans perte d'informations.

⇒ Réutilisation de l'information :

Une fois le document balisé l'information peut-être réutilisée par différentes applications. Ainsi on peut à partir d'un document balisé en SGML : le convertir en HTML (HyperText Markup Language), l'intégrer dans une base de données ou encore extraire la notice documentaire.

⇒ Cohérence et intégrité de l'information :

SGML élimine la traduction de données, donc il n'y a pas de perte d'information et, la structure qui est définie pour une classe de documents permet d'avoir une cohérence logique commune à la classe de documents.

2.3 LE LANGAGE SGML

SGML comprend un système de balisage descriptif. Contrairement au balisage procédural qui impose un traitement particulier du texte balisé, le balisage descriptif utilise des codes de balisage qui donnent des noms pour catégoriser les parties d'un document. Ces balises sont des identificateurs génériques. Par exemple, la balise <p> indique un début de paragraphe. Le balisage descriptif permet donc de décrire la structure logique d'un document.

2.3.1 La notation SGML

Pour différencier ce qui est du texte d'une balise, chaque identificateur générique est précédé par le symbole < s'il se trouve en début d'élément ou </ suivi du nom de l'identificateur et finit par >.

Ainsi le titre de ce rapport de stage pourra être balisé de la manière suivante :

<Titre> Intégration du langage SGML dans l'édition scientifique et technique </Titre>

2.4 LES COMPOSANTES DE SGML

2.4.1 La déclaration SGML

Commune à tous les documents elle permet de préciser le jeu de caractères à utiliser ainsi que leur codage (par exemple Isolat1 pour les caractères latins non accentués). La déclaration SGML permet également de définir les délimiteurs de balises qui vont être utilisés.

2.4.2 La DTD (Document Type Definition)

Une DTD définit les éléments, entités et attributs. Un élément est une partie logique d'un document, délimité par une balise de début et une balise de fin. L'entité est une unité

d'information à laquelle on se réfère avec un symbole dans une DTD ou une instance de document, les entités sont utilisées notamment pour les codages des caractères spéciaux. Enfin, les attributs sont les qualités caractéristiques d'un élément autre que le type ou le contenu.

La définition du type de document permet donc de définir les noms des éléments autorisés, leur fréquence d'apparition, leur ordre d'apparition, la présence obligatoire ou non des balises de début ou de fin, le contenu des éléments, les attributs des balises et le nom des entités qui doivent être utilisés.

La DTD est commune à une classe de documents. La norme ISO 12083 de 1994 définit des DTD pour les classes de documents suivantes :

- ⇒ les livres,
- ⇒ les articles,
- ⇒ les publications en série,
- ⇒ les formules mathématiques.

2.4.3 L'instance de document

Il s'agit du document balisé en SGML. Elle comporte le texte marqué par les balises SGML et la référence à la DTD qu'elle respecte.

2.5 LES OUTILS SGML

Pour créer des documents SGML, on peut utiliser les systèmes suivants :

- ⇒ un éditeur de texte courant,
- ⇒ un système de balisage automatique combiné avec un système de reconnaissance optique de caractères (OCR), pour une application rétrospective,
- ⇒ des feuilles de style type Word,
- ⇒ un éditeur SGML.

Les deux dernières possibilités sont celles qui ont le meilleur taux de réussite.

2.5.1 Les éditeurs SGML

SGML est un système de balisage descriptif, les DTD et instances de documents sont donc lisibles. On peut donc produire des documents SGML avec un simple éditeur de texte, SGML n'impose pas, comme nous l'avons vu précédemment de logiciels spécialisés. Néanmoins il existe des environnements spécifiques qui facilitent la saisie et l'édition de documents SGML.

Ces logiciels appelés éditeurs SGML peuvent avoir les caractéristiques suivantes :

- ⇒ reconnaissance des identificateurs génériques, attributs, entités définis dans n'importe quel document,
- ⇒ empêchent les erreurs de balisages,

- ⇒ possibilité d'extraire par exemple les grands titres pour faire une table des matières,
- ⇒ permettent d'indiquer le formatage du contenu des éléments,
- ⇒ conformité à la norme SGML,
- ⇒ peuvent être utilisés dans un mode où il n'y a pas de contrôle de conformité des documents,
- ⇒ sont capables d'ajouter des balises au document sans les afficher à l'écran.

2.5.2 Les parseurs

Il s'agit de logiciels qui permettent de s'assurer qu'un document SGML est cohérent et sans erreur. Ils vérifient que la DTD est conforme à la syntaxe SGML et que l'instance de document respecte la DTD.

Ils peuvent être utilisés indépendamment ou inclus dans un éditeur SGML. Il existe de nombreux parseurs : certains sont commercialisés et d'autres sont dans le domaine public.

Parmi les parseurs, on distingue les parseurs de validation SGML qui indiquent les erreurs SGML du document et les simples parseurs SGML qui ne reconnaissent que le balisage sans faire de contrôle.

2.6 L'INSTALLATION SGML

Après avoir défini ce qu'est SGML et les différents types d'outils qui existent, voici pour conclure le schéma d'une installation SGML type¹ :

¹ SGML Pratique / Eric Van Herwijnen.- Paris : Thomson publishing, 1995. p.29.

TROISIEME PARTIE

La méthodologie utilisée

Avant de commencer les analyses nécessaires pour la conception, il convient de définir ce qu'est un acte de colloque. Un acte de colloque est un document qui regroupe les textes des communications qui ont été présentées lors d'un colloque. Un acte de colloque est donc une collection de communications écrites présentées à un même colloque

A travers cette définition, on comprend bien qu'un acte de colloque est différent d'un livre. Il se rapproche plus d'une publication en série (collection d'articles) mais il n'a aucune périodicité. C'est pour cela qu'il n'est pas possible dans le cadre de ce projet d'utiliser les DTD normalisées et qu'il faut en créer une.

Cependant, pour concevoir la DTD on se basera sur les DTD normalisées (ISO 12083) car il est plus aisé de partir de ce qui existe et surtout il convient lors de la création de la DTD de respecter la norme pour avoir un système ouvert et ainsi, correspondre à l'esprit de la norme.

La méthodologie pour la conception de la DTD des actes de colloque se déroule en deux phases : tout d'abord l'analyse de l'existant et la définition des applications souhaitées et ensuite l'analyse des documents actes de colloque pour dégager l'ensemble des éléments qui seront dans la DTD.

1. Analyse préalable pour la conception de la DTD

1.1 DEFINITION DE LA CHAINE EDITORIALE

1.1.1 Les principes

En 1996, les scientifiques du Cemagref ont participé à 280 colloques.

Pour l'édition des actes de colloques, seuls sont édités par le Cemagref les colloques organisés ou coorganisés par le Cemagref. La parution des actes de colloque est généralement programmée pour le jour du colloque, néanmoins, l'édition peut se faire à posteriori.

L'ouvrage édité est le plus largement diffusé aussi bien en interne qu'en externe. Le volume moyen de tirage est de 500 exemplaires répartis de la manière suivante :

- ⇒ une grande partie est destinée aux participants du colloque,
- ⇒ 30 exemplaires à la division auteur pour une diffusion gratuite auprès de ses partenaires privilégiés,
- ⇒ 50 exemplaires pour la diffusion interne effectuée par la DICOVA,
- ⇒ le reste est destiné à la commercialisation qui est assurée par la DICOVA via son diffuseur.

1.1.2 Les acteurs et leur(s) rôle(s)

Pour l'édition des actes de colloque on distingue cinq acteurs : les auteurs des communications, le coordinateur ou l'éditeur scientifique, le comité scientifique du colloque, le comité d'organisation et le responsable d'édition.

◆ Les auteurs des communications :

Ils doivent fournir leur communication généralement trois mois avant la tenue du colloque. Le texte, relu et validé au niveau scientifique par un comité de lecture qui peut être le comité scientifique du colloque, est donné au responsable d'édition sous forme de fichier. En général, le logiciel utilisé est Word 6 sous Windows pour le texte, les illustrations viennent de toutes provenances : logiciel de traitement de texte, de dessin et sont fournies sur différents supports : papier, disquette, diapositive...

Le responsable d'édition fournit une note aux auteurs des communications. Celle-ci donne des indications sur la mise en forme du document mais aussi sur sa structuration notamment pour le début de la communication.

◆ Le coordinateur ou l'éditeur scientifique :

Après l'appel à communications, c'est lui qui collecte les textes et les transmet ensuite au responsable d'édition. Il peut aussi donner des consignes aux auteurs pour la rédaction de leurs communications : leur mise en page, leur lisibilité et leur homogénéité par rapport aux autres communications.

◆ Le comité scientifique du colloque :

Il a un rôle d'arbitre sur le sujet du colloque et a parfois pour mission d'organiser la relecture des manuscrits.

◆ Le comité d'organisation du colloque :

Chargé de tout ce qui concerne l'organisation matérielle du colloque, il n'intervient pas directement sur l'écriture ou l'édition des actes de colloque.

◆ Le responsable d'édition :

Travaillant au sein du service IST, il assure le suivi de l'édition du manuscrit et la fabrication du livre. Il structure l'ouvrage en fonction des thèmes abordés dans le colloque, il crée la table des matières et la couverture en respectant les contraintes données par le comité scientifique.

Les textes lui sont transmis dans un délai maximum de deux mois avant le colloque. Le délai d'impression des actes de colloque est de trois à quatre semaines.

1.2 LISTE DES APPLICATIONS ENVISAGEES

Afin de créer une DTD bien adaptée aux besoins, il est important de préciser clairement les applications qui sont envisagées :

- ⇒ une des premières applications est l'édition papier des actes de colloque,
- ⇒ la deuxième application est l'extraction de la notice documentaire pour le Cemagref et pour le programme Griseli,
- ⇒ la dernière application est la diffusion en ligne de tout ou partie des actes de colloque.

La mise en oeuvre de ces applications, sera développée dans la quatrième partie.

1.3 LES UTILISATEURS

Les premiers utilisateurs de la DTD seront les auteurs des communications puis le service IST qui sera chargé d'une part de l'édition papier des actes de colloque via l'entité édition et, d'autre part de l'extraction des notices documentaires pour les bases de données du Cemagref et pour le programme Griseli via l'entité gestion et diffusion de l'information.

2. Conception de la DTD

2.1 CONSTITUTION D'UN CORPUS D'ACTES DE COLLOQUE

Le corpus des actes de colloque comprend d'une part les actes de colloque édités par le Cemagref et d'autre part des actes de colloques édités par d'autres organismes au niveau national et international. Ce choix a été fait afin que la DTD ne soit pas trop fermée sur les habitudes du Cemagref.

Les actes de colloque ont été choisis en fonction de la cohérence de leur présentation et de leur diversité les uns par rapport aux autres.

En tout, 8 actes de colloque ont été choisis. Il s'agit de :

- ⇒ Les apports de la recherche scientifique à la sécurité neige, glace et avalanche. Chamonix : 30 mai - 3 juin 1996. Antony : Cemagref éditions, 1995.
- ⇒ *Acipenser : actes du premier colloque international sur l'esturgeon*. Bordeaux : 3-6 octobre 1989.- Antony : Cemagref éditions, 1991.
- ⇒ *Measures for success : metrology and instrumentation in aquaculture management*. Bordeaux : 23-25 mars 1994.- Antony : Cemagref éditions, 1994.
- ⇒ *Hydrotop 94 : colloque "mieux gérer l'eau"*.- Marseille : 12-15 avril 1994.
- ⇒ *Ageng 94 International Conference on Agricultural Engineering*.- Milan : 29 août- 1 septembre.
- ⇒ *Toward a practice of autonomous systems : proceedings of the First European Conference on Artificial life*, edited by Fransisco J. Varela and Paul Bourguine.- Paris : décembre 1991.
- ⇒ *Autonomie et interactions fonctionnelles : actes des journées de Rochebrune*. Rochebrune : 17-21 janvier 1994.- Paris : France Télécom CNET, 1994.
- ⇒ *15th congress on irrigation and drainage : water management in the next century*. La Haye : 4-5 septembre 1993. Coordinateur Emile Lorre (Cemagref).- Paris : ICID-CID, 1993.

2.2 ANALYSE DU CORPUS ET REPERAGE DE LA STRUCTURE LOGIQUE

Avant de commencer l'analyse, il faut trouver un nom à la DTD. Ce nom ne doit pas dépasser huit caractères. La DTD qui va être créée sera appelée DTD Congrès.

Afin de correspondre aux mieux aux besoins qui ont été exprimés, cette analyse s'est faite en étroite collaboration avec le service édition en tenant compte des applications voulues avec la

DTD. Selon les applications souhaitées les implications dans la structure logique ne sont pas les mêmes. En effet, pour que la DTD puisse être compatible avec celle de Griseli, il faut veiller à reprendre les mêmes éléments que ceux présents dans la DTD de l'INIST.

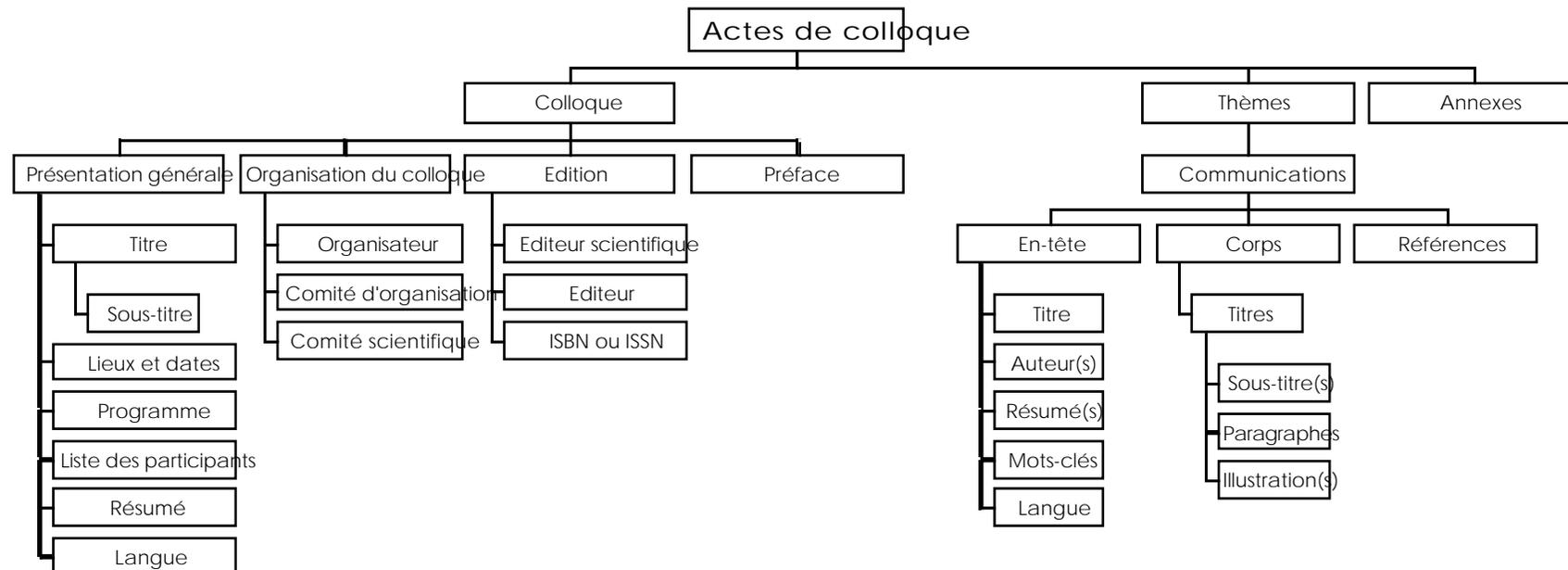
Comme nous l'avons vu ci-dessus, un acte de colloque se décompose en deux parties : d'une part l'acte de colloque en général et, d'autre part les communications. Ces deux choses ont d'abord été analysées séparément. De plus, dans cette analyse nous avons fait attention à ne pas être influencé par la structure physique.

Dans un premier temps, nous avons listé les principaux éléments existants, leur contenu et les propriétés qu'ils entretiennent entre eux. Par exemple, le titre, le nom de l'éditeur sont des éléments logiques, en revanche la table des matières n'est pas un élément logique puisqu'elle peut être générée automatiquement à partir des thèmes des communications. Tous les éléments susceptibles d'apparaître dans l'édition d'un acte de colloque ont été listés. L'objectif est de prévoir toutes les situations possibles. Par exemple, il faut prévoir qu'un acte de colloque peut être édité dans une série et donc possède un ISBN et un ISSN.

Dans un deuxième temps, nous avons hiérarchisé les éléments, cette étape a abouti à la conception d'un schéma arborescent qui permet de visualiser la structure logique des actes de colloque.

Ce schéma ne détaille pas tous les éléments, il représente seulement les éléments susceptibles d'apparaître dans l'édition d'un acte de colloque. Il n'indique ni l'obligation de présence ou non des éléments, ni le nombre de fois où ils sont susceptibles d'apparaître ni l'ordre d'apparition des éléments.

La structure logique des actes de colloque



En ce qui concerne l'élément organisation du colloque, nous avons distingué l'organisateur (personne morale qui prend en charge la responsabilité d'organisation du colloque), le comité d'organisation et le comité scientifique. Nous avons remarqué dans certains actes de colloque la présence de comité international d'organisation, nous l'avons inclus avec le comité d'organisation.

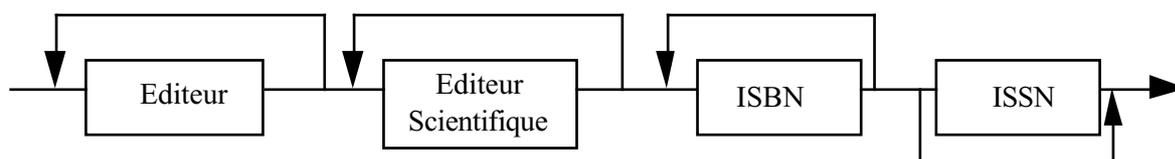
Certains éléments ont été rajoutés car ils nous ont paru indispensables, c'est le cas du résumé et de la langue présents à la fois dans l'élément colloque et dans l'élément communication. Ces éléments sont utiles pour la notice documentaire du Cemagref et du projet Griseli.

2.3 LES PRINCIPAUX ELEMENTS DE LA DTD CONGRES

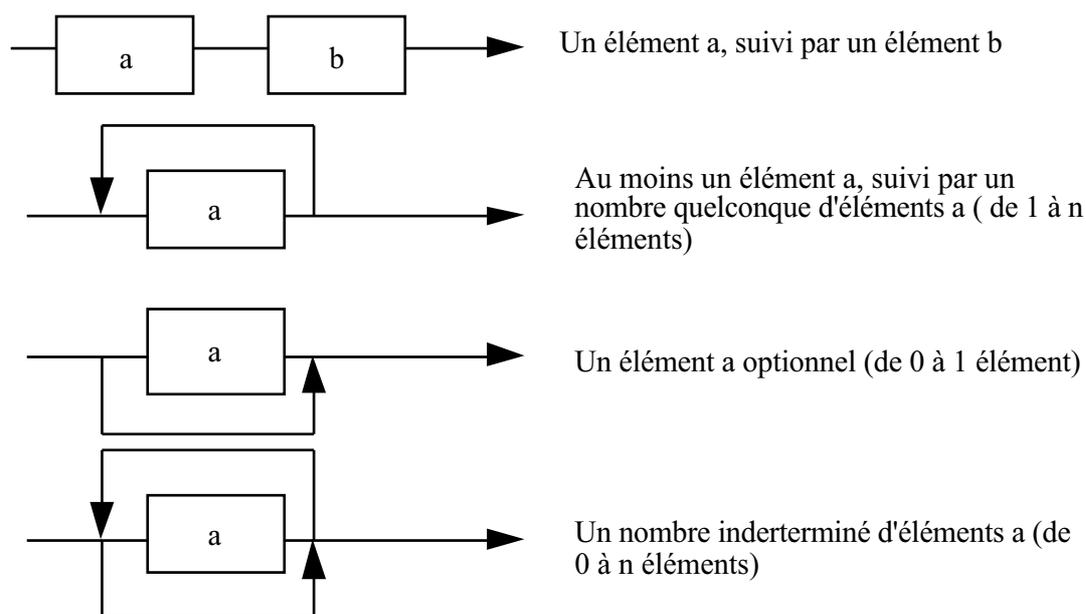
Dans cette étape, il s'agit de reprendre les éléments qui ont été définis ci-dessus et de choisir s'ils sont des éléments, des entités ou des attributs. Pour ce faire il faut se poser certaines questions telles que : les données nécessitent-elles un traitement particulier ? Au quel cas il convient de les séparer et d'en faire un élément à part entière. Les données sont-elles décrites dans le contenu ?

Pour définir les principaux éléments de la DTD congrès, nous avons utilisé les diagrammes de structure qui nous permettent d'enrichir le schéma arborescent. En effet, ils indiquent l'ordre d'apparition des éléments, leur nombre d'occurrences ainsi que l'obligation de présence ou non de ces éléments. De plus, ces diagrammes de structure peuvent être, par la suite, traduits facilement en déclaration de balisage SGML.

Chaque diagramme de structure reprend un noeud de l'arbre. A titre d'exemple voici le diagramme de structure du noeud de l'élément édition de l'acte de colloque ainsi que la légende. Les autres diagrammes sont fournis en annexe.



Légende :



Après avoir réalisé les diagrammes de structure, nous avons étudié attentivement la norme ISO 12083 en nous référant à la fois aux DTD livres, périodiques et articles. En effet, la norme définit un nombre important d'éléments qu'il convient de reprendre afin d'avoir une compatibilité maximum entre les DTD mais aussi parce qu'il est plus facile de partir de quelque chose qui existe que de recréer tout. C'est pour cela que les noms des éléments repris de la norme sont en anglais, nous ne les avons pas traduits en français. De plus à partir de la consultation de la norme nous avons regroupé certains éléments, ainsi tout ce qui concerne l'organisation du colloque et le programme et la liste des participants ont été regroupés dans un même élément nommé groupe de travail. Le tableau suivant définit les noms des éléments (d'après la norme ISO 12083) de la DTD Congrès et précise s'ils sont obligatoires, répétables ou non.

Hiérarchie et Balises	Eléments	OBLIGATOIRE / FACULTATIF	REPETITIVITE
Congrès	Racine de la DTD	0	×
0. FRONT	Préliminaire du colloque	O	×
0.1 Titlegrp	Groupe titres	O	×
0.11 Title	Titre du colloque	O	R
0.1.2 Subtitle	Sous-titre	F	R

0.2 Authgrp	Groupe auteurs	O	×
0.2.1 Author	Auteur	F	R
0.2.2 Corpauth	Collectivité auteur	F	R
0.2.3 Aff	Affiliation	F	R
0.3 Pubfront	Éléments de la publication	O	×
0.3.1 Date	Date	O	×
0.3.2 Pubname	Nom de l'éditeur	O	R
0.3.3 Location	Lieux d'édition	F	R
0.3.4 Confgrp	Groupe Congrès	O	×
0.3.4.1 NO	Numéro	F	×
0.3.4.2 Confname	Nom du congrès	O	×
0.3.4.3 Date	Date	O	×
0.3.4.4 Location	Lieux	O	×
0.3.4.5 Sponsor	Organisateur	F	R
0.3.5 ISBN	ISBN	F	R
0.3.6 ISSN	ISSN	F	×
0.3.7 Extent	Nombre de pages	O	×
0.4 Abstract	Résumé	O	R
0.5 Language	Langue	O	R
0.6 Grptravail	Informations sur le congrès	F	×
0.6.1 Intitulé	Intitulé du groupe de travail (comité d'organisation, scientifique ou liste des participants)	O	R
0.6.2 Authgrp	Liste de noms des	O	R

	comités		
0.7 Programme	Programme	F	×
0.8 Préface	Préface de l'acte	F	R
1. BODY	Corps de l'acte de colloque	O	×
1.1 Thème	Thème des communications	O	R
1.1.1 Sous-thème	Sous-thème des communications	F	R
1.1.2 Communication	Communication	O	R
<i>1.1.2.1 Comfront</i>	Préliminaire de la communication	O	×
1.1.2.1.1 Titlegrp	Groupe titre	O	×
1.1.2.1.1.1 Title	Titre	O	R
1.1.2.1.1.2 Subtitle	Sous-titre	F	R
1.1.2.1.2 Authgrp	Groupe auteur	O	×
1.1.2.1.2.1 Author	Auteur	F	R
1.1.2.1.2.2 Corpauth	Collectivité auteur	F	R
1.1.2.1.2.3 Aff	affiliation	F	R
1.1.2.1.3 Abstract	Résumé	O	R
1.1.2.1.4 Keyword	Mot-clé	F	R
1.1.2.1.5 Language	Langue	O	×
<i>1.1.2.2 Combody</i>	Corps de la communication	O	×
1.1.2.2.2 Part	Partie de la communication	O	R
1.1.2.2.2.1 Chapter	Chapitre	F	R
1.1.2.2.2.1.1 Section	Section	F	R

1.1.2.2.2.1.1.1 Subsect	Sous-section	F	R
1.1.2.3 <i>Combiblist</i>	Références bibliographiques	F	R
2. Annexes	Annexes de l'acte de colloque	F	R

A ce stade, nous possédons tous les éléments de la DTD Congrès. Avant de passer à l'écriture de la DTD, nous avons trouvé indispensable de revenir aux notices documentaires du Cemagref et du projet Griseli pour vérifier qu'il ne manquait pas d'éléments pour pouvoir par la suite établir des correspondances.

2.4 COMPARAISON DE LA DTD CONGRES AVEC LA DTD GRISELI ET LA NOTICE DOCUMENTAIRE DU CEMAGREF

Dans cette comparaison, nous avons distingué les notices qui concernent les actes de colloque dans leur globalité et les notices se rapportant à une communication dans un acte de colloque.

Le but n'est pas que la DTD Congrès reprenne tous les éléments des notices du Cemagref et du projet Griseli mais en priorité les éléments qui obligatoires dans la DTD Griseli et dans la notice documentaire du Cemagref.

Un tableau comparatif de ces éléments est fourni en annexe, il indique pour chaque champ s'il existe une correspondance ou non.

2.4.1 Comparaison avec la notice du Cemagref

Pour établir la comparaison entre la DTD Congrès et la notice documentaire du Cemagref, nous nous sommes appuyé sur le format de notice documentaire défini par le Cemagref.

Nous avons constaté que les éléments suivants de la notice documentaire ne font pas partie de la DTD Congrès :

- ⇒ Type de document
- ⇒ Sigle
- ⇒ Pages : page de début et de fin d'une communication
- ⇒ Coll : collection ou série
- ⇒ Localis : localisation du document
- ⇒ Descripteur : appartenant au thesaurus du Cemagref

Les champs type et localisation peuvent être incrémentés automatiquement. Les champs descripteur, sigle, coll devront être remplis par un documentaliste du Cemagref.

Concernant le champ page, il ne nous paraît pas utile de le rajouter comme élément dans la DTD. En effet, la pagination dépend du support sur lequel est éditée la communication. C'est donc plus un élément physique que logique.

2.4.2 Comparaison avec la DTD Griseli

Pour la comparaison, nous avons repris un à un les éléments définis dans la DTD Griseli. Après étude, nous constatons que les éléments suivants ne font pas partie de la DTD Congrès :

- ⇒ Le sigle de l'organisme fournissant la notice,
- ⇒ la date de création de la notice,
- ⇒ le numéro d'origine de la notice,
- ⇒ le type de document,
- ⇒ la nature du document,
- ⇒ code de classement de l'INIST,
- ⇒ mots-clés issus du thesaurus Pascal et/ ou du thesaurus du Cemagref,
- ⇒ numéro de page de la communication
- ⇒ tous les éléments concernant l'accès au document.

Tout ce qui concerne l'accès au document peut être incrémenté de manière automatique par le service de documentation du Cemagref ou par l'INIST.

Tout comme pour la notice du Cemagref, les champs : sigle de l'organisme fournissant la notice, date de création, numéro d'origine, type de document, la nature du document, pourront être incrémentés automatiquement.

En revanche les champs restants devront être remplis par un documentaliste du Cemagref.

Dans l'ensemble que ce soit pour les notices documentaires du Cemagref ou pour les notices du projet Griseli les principaux éléments sont présents dans la DTD Congrès.

3. *Ecriture de la DTD*

Afin de pouvoir comprendre l'écriture de la DTD Congrès, nous allons présenter les principales règles d'écriture du langage SGML.

3.1 REGLES D'ECRITURES

Une DTD se décompose en quatre parties :

- ⇒ la déclaration de type de document,
- ⇒ la déclaration d'entités,
- ⇒ la déclaration d'éléments,
- ⇒ la déclaration d'attributs.

La structure d'une déclaration de balisage contient l'un des mots-clés suivants : ENTITY, ELEMENT, ATTLIST suivi du nom et du contenu.

Ces déclarations peuvent contenir des connecteurs et/ou des indicateurs d'occurrence qui indiquent les relations qui existent entre les éléments.

Les connecteurs sont au nombre de trois :

- ⇒ & : connecteur et
- ⇒ , : connecteur de séquence, exemple : a, b signifie a suivi de b
- ⇒ | : connecteur ou

Les indicateurs d'occurrence sont également au nombre de trois :

- ⇒ ? : indicateur d'occurrence optionnelle 0 ou 1,
- ⇒ + : indicateur d'occurrence obligatoire 1 fois ou plus,
- ⇒ * : indicateur d'occurrence obligatoire 0 ou plus.

A noter que les indicateurs d'occurrence ont une priorité supérieure aux connecteurs.

Exemple de déclaration pour l'élément groupe titre :

```
<! ELEMENT titlegrp - - (title, subtitle)* >
```

Les caractères "- -" précisent si on peut omettre ou non les balises de début ou de fin de l'élément. Par prudence, nous avons interdit dans la DTD Congrès la minimisation par omission de balise.

Enfin, il est conseillé de commenter la DTD afin de la rendre plus lisible. Les commentaires sont encadrés des caractères : "--".

3.2 ECRITURE ET VERIFICATION

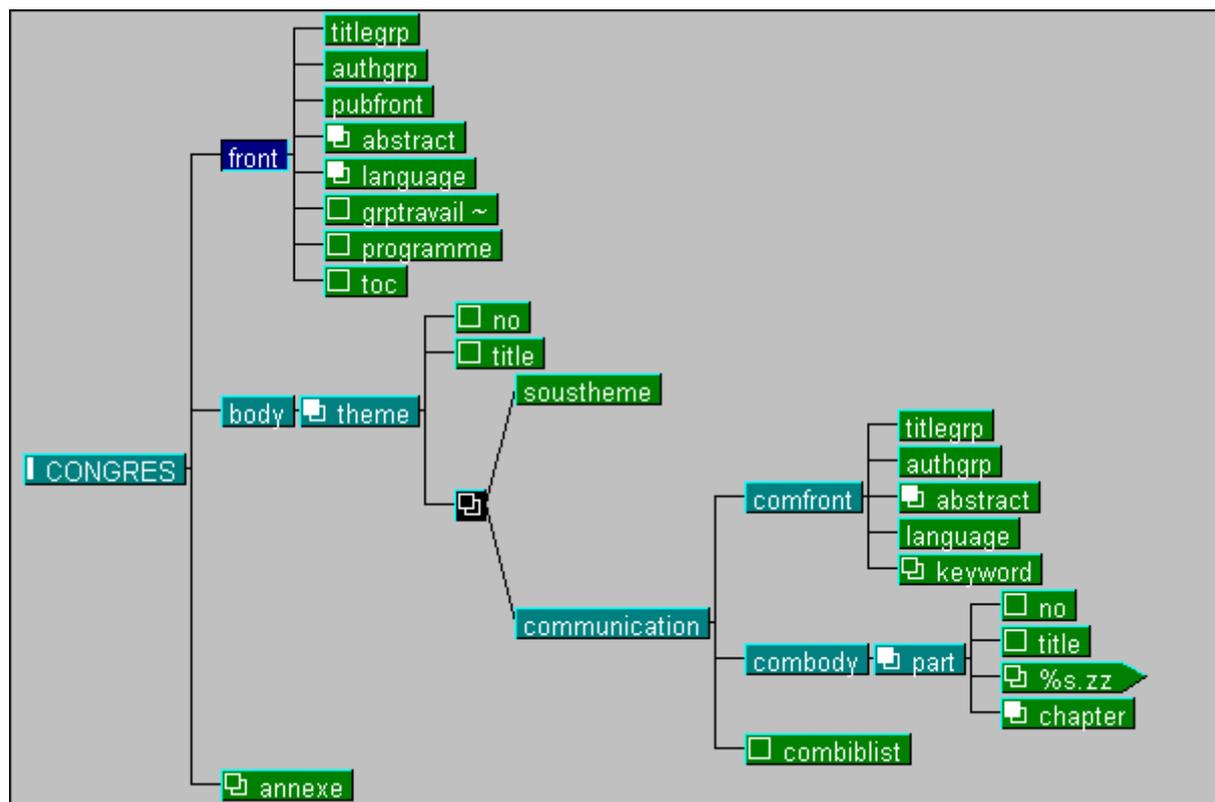
La DTD Congrès, que vous trouverez ci-jointe en annexe, a été écrite sous Word en format texte. Pour l'écriture nous avons, chaque fois que cela était possible, repris des DTD ISO 12083 et Griseli les éléments dont nous avons besoin.

Une fois la DTD écrite, nous nous sommes servis de deux logiciels :

- ◆ **Near and Far Lite** qui permet de visualiser la structure arborescente de la DTD et de signaler les erreurs existantes.

La vue arborescente de la DTD nous permet de mieux repérer les oublis qui ont pu être faits ou les erreurs logiques de construction.

Exemple d'écran avec Near&Far Lite :



- ◆ **NSGMLS** : c'est le plus récent des parseurs du domaine public. Il fonctionne sous DOS.

Il indique toutes les erreurs commises dans la DTD : erreur de syntaxe, erreur dans les déclarations...

Exemple de message d'erreur :

dtdacte2.dtd:77:1:E: character `<' invalid: only delimiter `>', inclusions, exclusions and parameter separators allowed

Ligne 77 : <! ELEMENT %doctype; - - (front,body,annexe*)

Il manque la balise de fermeture de la déclaration d'élément.

Le parseur comme nous l'avons vu dans la deuxième partie permet de valider la DTD. Cependant même si la DTD est validée cela ne suffit pas, il faut ensuite faire des instances de documents pour vérifier si tout est correct et qu'aucun élément n'a été oublié. Dans cette DTD, nous avons pris le parti de faire quelque chose de très simple aussi nous avons réduit les déclarations d'attributs.

QUATRIEME PARTIE

*Intégration de la DTD dans
la chaîne éditoriale du Cemagref*

Dans cette partie nous allons rappeler le schéma de la chaîne éditoriale "classique" puis nous présenterons les différents éditeurs SGML disponibles sur le marché.

Ensuite, nous proposerons les différentes possibilités d'intégration de SGML dans la chaîne éditoriale du Cemagref, les contraintes existantes pour chaque possibilité et les problèmes matériels et humains qui en découlent.

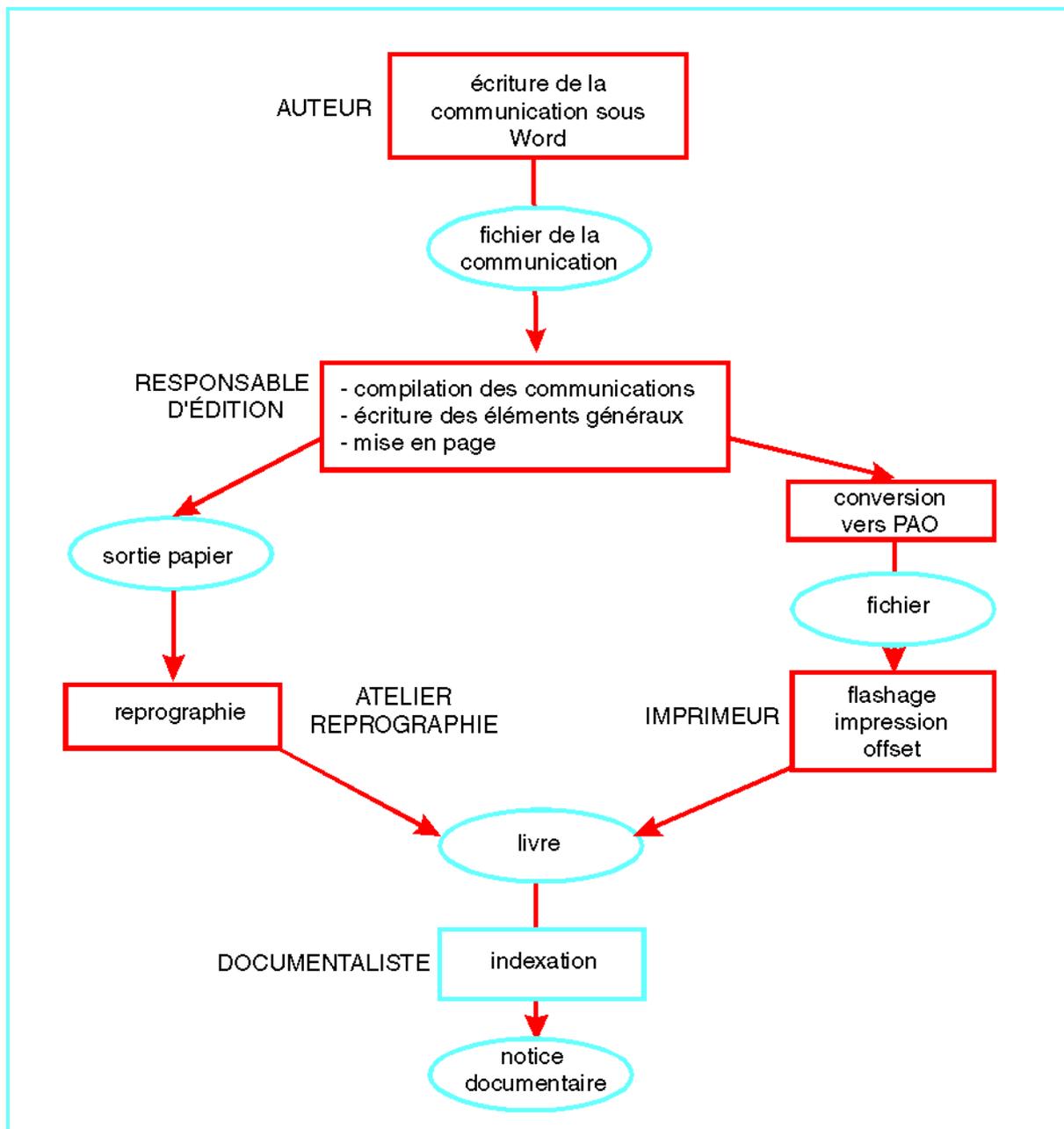
Cependant avant d'intégrer SGML dans la chaîne éditoriale, il faut savoir que la DTD Congrès devra être affinée c'est-à-dire créer des éléments pour les notes de bas de page, les équations, les figures ou illustrations ainsi que les listes. Ce travail peut être fait environ en quinze jours.

Pour l'intégration de SGML dans la chaîne éditoriale, nous allons distinguer trois maillons : la production de texte, l'impression des actes de colloque, la diffusion en ligne et l'extraction des notices.

Enfin, dans la dernière partie nous présenterons le schéma de la chaîne éditoriale avec l'intégration de SGML et les conclusions de la mission.

1. SGML et la chaîne éditoriale

Avant de présenter les diverses solutions pour intégrer le langage SGML dans la chaîne éditoriale, il nous a paru important de rappeler le schéma de la chaîne actuelle pour les actes de colloques.



Dans la chaîne éditoriale actuelle, on se rend bien compte qu'une grande partie du travail se fait manuellement. Ainsi, le responsable d'édition doit souvent remettre en page les communications qui lui sont fournies. En effet, les auteurs ne respectent pas toujours la note aux auteurs remise par le responsable d'édition. Les notices documentaires sont également faites manuellement par les documentalistes du Cemagref.

Dans ce schéma, on a distingué deux sortes d'impressions. En effet, dès l'instant où l'acte de colloque est imprimé à plus de cinq cents exemplaires on a besoin d'une impression offset. Pour la réaliser, on doit auparavant faire un film d'après le fichier du document. Or, les fichiers Word ne peuvent être flashés, il faut donc faire une conversion vers un logiciel de PAO.

Pour les tirages de moindre importance, la technique utilisée est la reprographie qui s'effectue à partir d'une sortie papier. La reprographie du document peut s'effectuer au Cemagref qui dispose d'un atelier d'imprimerie.

2. La production de texte

Pour la production de texte, c'est-à-dire la création d'instance, nous avons besoin d'un éditeur SGML. Nous allons donc avant de présenter les différentes solutions d'intégration de SGML, dresser un panorama des éditeurs SGML disponibles sur le marché.

2.1 LES EDITEURS SGML

En France, l'utilisation du langage SGML est encore assez marginale, nous avons donc rencontré des difficultés pour avoir des informations pertinentes sur ces logiciels et, aucune version de démonstration ne nous a été fournie. Nous avons acheté deux logiciels, il s'agit de Near&Far Author et de SGML Author. Cependant, le manque de temps ne nous a pas permis de tester complètement ces produits.

L'application souhaitée par le Cemagref doit rester quelque chose de simple. Aussi nous avons décidé de nous attacher exclusivement aux logiciels qui fonctionnent avec Word, d'autant plus que ce logiciel est utilisé par les auteurs des communications. Il s'agit donc d'éditeurs qui fonctionnent comme des utilitaires dans un logiciel de traitement de texte.

	Near&Far Author 1.1	SGML Author 1.0	SoftQuad Author/Editor 3.5
Producteur	Microstar	Microsoft	SofQuad
Fonctionnalité	<p>Interface graphique : visualisation de la structure arborescente de la DTD (voir la copie d'écran de Near&Far Lite),</p> <p>Visualisation des liens qui existent entre les éléments de la DTD et les styles de Word,</p> <p>Une boîte de dialogue présente automatiquement le futur</p>	<p>Ce logiciel se décompose de la manière suivante :</p> <p>SGML Author qui permet de créer un fichier associé qui comprend les relations qui existent entre Word et les éléments SGML,</p> <p>Un convertisseur, Des feuilles de style Word pour les auteurs.</p>	<p>Interface graphique, Logiciel Wysiwyg, Création de modèle pour l'écriture de document SGML, Supporte les graphiques, la vidéo et le son,</p> <p>Possibilité de voir différentes vues du document.</p>

	élément que l'on doit insérer et lui applique le style qui a été défini par l'utilisateur.		
Présence d'un parseur	Oui	Oui	Oui
Configuration minimum	486/66 PC 12 MB RAM, 16 recommandée Microsoft Word 6.0 Windows 3.1	466 PC 6 MB RAM, 8 recommandée Microsoft Word 6.0 Windows 3.1	486 PC Windows 3.1
Prix hors-taxe du logiciel	2 200,00 F	6 600,00 F	995,00 F

Avec son interface graphique, Near&Far Author permet à l'éditeur de créer facilement des styles : l'éditeur travaille avec la structure arborescente et pour chaque élément il associe un style. Ensuite, il fait valider la création du style par le logiciel. Pour l'auteur qui crée une instance, une boîte de dialogue apparaît à l'écran indiquant les éléments qu'il doit insérer dans le texte. En parallèle il voit la structure arborescente de la DTD et, en surbrillance, l'endroit où il se trouve dans la DTD. Ceci peut éventuellement troubler l'auteur.

Pour SGML Author, nous avons pu remarquer qu'il est plus destiné aux auteurs. En effet, la création d'instance est aisée : une barre d'outil apparaît avec dessus le nom des éléments que l'auteur doit insérer dans le texte. Par contre, la création des styles se fait dans un module indépendant de Word, on n'a donc pas la possibilité de voir directement le résultat de la création des styles.

Concernant le logiciel SoftQuad Author/Editor nous n'avons pas pu avoir de version de démonstration et donc évaluer correctement le produit. Par contre, de part les informations que nous avons pu avoir, notamment sur Internet, ce logiciel ne semble pas être beaucoup utilisé.

Notre choix s'est donc porté sur Near&Far Author. Même s'il n'est pas forcément simple pour les auteurs, il est le logiciel le plus facile d'utilisation et le plus performant. En effet, avec SGML Author, la création des styles nous a paru beaucoup trop complexe. De plus, Near&Far Author est un logiciel distribué par AIS/Berger Levrault filiale de Berger Levrault spécialisée dans l'édition électronique. On peut donc compter sur une maintenance compétente. Enfin, Near&Far Author semble être le logiciel qui a le plus de chances d'évoluer.

Quelle que soit la solution choisie pour l'intégration de SGML dans la production de texte, il est préférable de choisir le même logiciel pour le responsable d'édition et pour les auteurs ce qui élimine les problèmes de conversion et de compatibilité.

2.2 LES DIFFERENTES SOLUTIONS D'INTEGRATION DE SGML

Pour la création d'instances, deux scénarios sont envisageables :

◆ Premier scénario :

L'auteur balise en SGML le texte de sa communication et le responsable d'édition compile les communications et fait la présentation générale du colloque en SGML.

Ce scénario nécessite une très bonne formation des auteurs. Cette formation pourra être faite par le biais d'une note aux auteurs et d'un guide d'utilisation du logiciel choisi. De plus, il faut prévoir un interlocuteur permanent appartenant à l'entité édition avec qui l'auteur pourra entrer en contact s'il rencontre des problèmes dans le balisage de sa communication.

Dans ce scénario, l'éditeur devra avant toute chose posséder de bonnes connaissances en SGML et avoir une parfaite maîtrise du logiciel qui sera fourni aux auteurs. De plus avant de fournir aux auteurs le logiciel, le responsable d'édition devra pour chaque élément définir le style qu'il souhaite que les auteurs utilisent. Pour acquérir ces connaissances, le responsable d'édition pourra par exemple participer à une formation donnée par le producteur du logiciel. Cet apprentissage nécessite une durée de un à deux mois.

L'avantage de cette solution est qu'elle implique les auteurs dès la phase pré-éditoriale. Cependant il faut prévoir l'achat de nombreuses licences d'éditeurs SGML pour les auteurs et le logiciel utilisé doit être très convivial et simple d'utilisation. Un autre inconvénient est la diversité des sources de balisage qui peut augmenter le nombre d'erreurs. De plus, ce scénario nécessite une coopération totale des auteurs.

◆ Deuxième scénario :

L'auteur ne balise pas sa communication en SGML, il entre son texte en suivant une feuille de style Word. Le responsable d'édition serait donc chargé de convertir les communications en SGML et de faire la présentation générale du colloque.

Il faudrait donc fournir une feuille de style aux auteurs faites à partir d'un éditeur SGML et, ensuite convertir les fichiers Word en fichiers SGML. Near&Far Author permet d'associer à la DTD une feuille de style Word, prédéfinie par le responsable d'édition. Ensuite, le document Word respectant la feuille de style peut-être converti assez aisément en SGML.

Pour ce scénario, il faudra, comme pour le scénario précédent, créer une note aux auteurs qui présente et définit la feuille de style que le responsable d'édition fournirait aux auteurs, et prévoir une formation pour le responsable d'édition.

L'avantage de cette solution est que le balisage SGML est fait par une seule et même personne : le responsable d'édition. Ceci réduit les risques d'erreurs à la condition que les auteurs respectent bien la feuille de style.

2.3 QUEL CHOIX ?

Parmi ces deux solutions, la dernière semble être la plus réalisable. En effet, le premier scénario demande une forte coopération des auteurs.

L'investissement matériel de la dernière solution est assez faible : achat d'une ou deux licences du logiciel. L'investissement humain est quant à lui difficile à chiffrer. Il faut cependant compter au moins deux mois d'apprentissage du logiciel et du langage SGML pour le responsable d'édition et, un mois pour la création de la feuille de style et, la rédaction d'une note aux auteurs. Cette période comprend également une phase d'expérimentation.

Pour la réalisation de ce scénario, il serait peut-être préférable de scinder la DTD Congrès en deux DTD : une DTD concernant la présentation générale de l'acte de colloque destinée au responsable d'édition et une autre DTD pour les communications destinée aux auteurs.

En effet, en ayant deux DTD, la conversion des communications en SGML peut poser moins de problèmes. Avec une seule DTD, il faut convertir et ensuite intégrer la communication dans l'instance de l'acte de Congrès.

Pour scinder la DTD Congrès en deux, on pourra s'appuyer sur les DTD normalisées des périodiques et des articles. Dans la DTD périodique, on fait appel à la DTD article.

Il conviendra d'expérimenter les deux solutions et comparer l'usage pour connaître la solution qui pose le moins de problèmes.

L'intégration de SGML pour la production de texte devra se faire dans un premier temps avec un groupe restreint d'auteurs avant d'étendre cette solution à tous les auteurs.

Pour l'intégration de SGML dans la production des textes, il faut prévoir :

- ⇒ L'achat d'une licence Near&Far Author,
- ⇒ Une formation du responsable d'édition,
- ⇒ La création d'une feuille de style attachée à la DTD Congrès,
- ⇒ La création d'une note aux auteurs.

2. L'impression

A ce stade, on possède l'instance de l'acte de colloque et on veut faire une sortie papier du document mis en forme.

Pour proposer des solutions, nous allons distinguer deux possibilités : l'impression d'actes de colloque avec un tirage supérieur à 500 exemplaires et, l'impression avec un tirage moindre. En effet, selon le nombre d'exemplaires imprimés, les moyens mis en oeuvre ne sont pas les mêmes, comme nous avons pu le voir en début de quatrième partie.

◆ Les tirages de plus de 500 exemplaires:

Pour pouvoir effectuer un tirage important d'un acte de colloque, le responsable d'édition doit fournir un fichier du document mis en forme avec un logiciel de PAO (Publication Assistée par Ordinateur). En effet, à partir de ce fichier, l'imprimeur va pouvoir faire un film du document (cette opération s'appelle le flashage) qui servira ensuite pour l'impression.

Word n'est pas un logiciel de PAO. Le responsable d'édition devra donc transposer le fichier Word dans un logiciel de PAO. Ce dernier devra être capable d'appliquer des styles en fonction des balises.

Une autre possibilité est de donner à un imprimeur spécialisé l'instance de l'acte de colloque. Nous avons pris contact avec quelques imprimeurs pour connaître les avantages et inconvénient d'une impression à partir d'une instance de document.

Il en ressort que pour le Cemagref, l'impression des actes de colloque à partir d'un fichier SGML n'est pas une solution à envisager. En effet, les imprimeurs qui utilisent SGML le font car en même temps ils proposent un service complet, c'est à dire la création de la DTD et son intégration dans la chaîne éditoriale. Or, pour le Cemagref, la situation à envisager est celle où la structure SGML existe déjà.

◆ Les tirages de moins de 500 exemplaires:

Les petits tirages nécessitent une sortie imprimée du document pour ensuite pouvoir reprographier. Cette sortie doit évidemment être mise en forme. Le logiciel choisi ci-dessus : Near&Far Author utilise les ressources de Word et permet donc d'appliquer des styles à l'instance et de mettre en forme l'acte de colloque.

Dans le cas de petits tirages, l'intégration de SGML ne pose donc aucun problème, à partir de l'instance de l'acte de colloque, le responsable peut faire une sortie papier du document mis en forme.

3. La diffusion en ligne

La diffusion en ligne consiste à convertir l'instance de l'acte de colloque en HTML. HTML est un langage qui dérive de SGML, il met en oeuvre des liens hypertextes. Ce langage est utilisé pour la création de pages Web sur Internet. La conversion de SGML vers HTML ne devrait donc pas poser de problème.

En concertation avec le service IST, nous avons décidé que seule la table des matières serait au format HTML, les communications seraient fournies en format TXT, RTF ou PDF dans des fichiers séparés. Ces formats sont des formats d'échange fréquemment utilisés sur Internet.

Le format TXT est un format assez pauvre car ne permettant de présenter que du texte (figures et illustrations sont exclues), le problème est donc que nous allons perdre tous les dessins et figures qui sont dans les communications. Avec RTF (Rich Text Format), qui est un format propriétaire de Microsoft, nous conservons la mise en page des documents. Cependant ces fichiers ne peuvent se lire qu'avec un logiciel Microsoft. Enfin, le format PDF (Portable Document Format) est un format de la société Adobe dont l'objectif est de permettre la lecture de tout type de documents à partir d'un logiciel unique : Acrobat Reader d'Adobe. Ce logiciel est distribué gratuitement. Il dispose notamment d'un tourne-page pour faciliter la lecture des documents.

Sur ces trois formats, PDF est sans doute le plus adapté pour diffuser les textes intégraux des communications du Cemagref qui sont souvent très illustrées.

4. l'extraction des notices documentaires

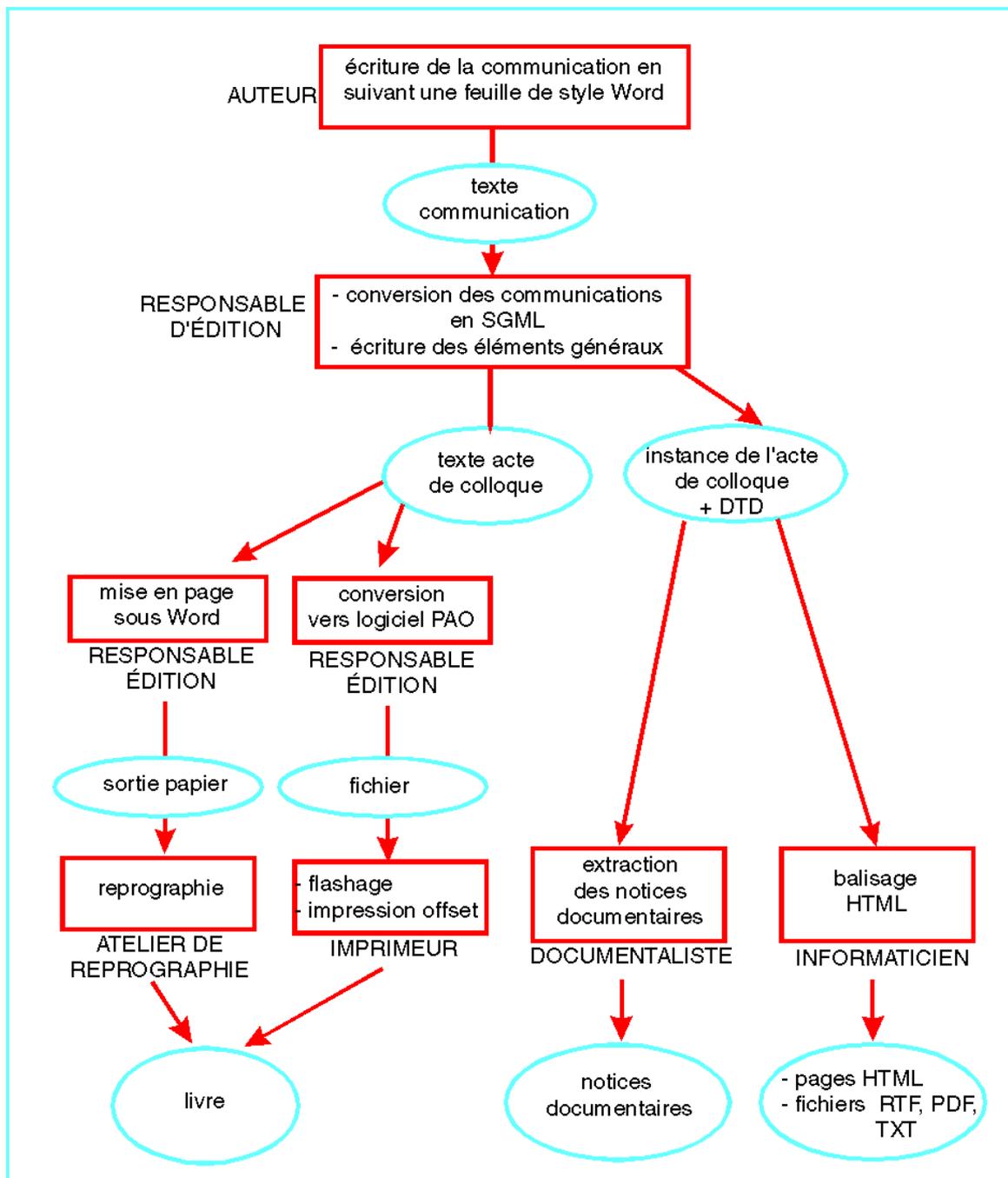
Pour extraire les notices (du colloque et des communications) destinées aux bases de données du Cemagref, il faut écrire un programme qui permette à partir de l'instance du document d'extraire l'information. Dans ce programme on indiquera dans quelle balise, il faut prendre l'information en indiquant ensuite dans quel champ de la notice cette information doit être rebasculée. Pour faire ce programme, on s'aidera du tableau comparatif qui a été fait entre la notice du Cemagref et la DTD Congrès. Pour l'écriture de ce programme, il faut compter deux à trois jours. Le langage de programmation utilisé pourra être le basic.

Pour l'extraction des notices documentaires destinées au projet Griseli, il suffit d'utiliser les macros Word qui ont été écrites avant par le Cemagref pour la réalisation de la coopération entre l'INIST et le Cemagref. Ces macros permettent à partir d'une notice du Cemagref de la convertir en SGML en respectant la DTD Griseli.

L'extraction des notices documentaires devra donc se faire en deux étapes : d'abord les notices pour les bases de données du Cemagref et ensuite les notices pour le projet Griseli.

5. SGML intégré dans la chaîne éditoriale

Pour mieux comprendre comment va se dérouler l'intégration de SGML dans la chaîne éditoriale du Cemagref, nous allons présenter le schéma de cette intégration avec les solutions qui ont été proposées auparavant.



6. *Eléments de réponse aux questions posées dans le cadre de la mission*

Nous allons dresser, sous forme de tableau, dans cette partie le bilan de la mission par rapport aux objectifs que nous nous étions fixés.

OBJECTIFS	RESULTATS
<p>Démontrer la maniabilité d'une application SGML dans le cadre de l'édition scientifique et technique, étudier les problèmes de mises en oeuvre (y compris les problèmes de coût au niveau matériel et humain) que cela implique et identifier les contraintes.</p>	<p>Oui, SGML peut s'appliquer dans l'édition scientifique et technique. Ce langage est assez maniable pour pouvoir s'adapter à des situations différentes telles que dans le cas présent : l'édition d'actes de colloque plus ou moins structurés avec un tirage faible. SGML n'est donc pas un langage réservé aux seuls gros projets d'édition. Pour les coûts, au niveau matériels, ils sont assez réduits par contre les coûts humains sont assez lourds pour la mise en oeuvre du projet.</p>
<p>Démontrer si cela améliore la productivité pour l'édition des actes de colloque.</p>	<p>Il est difficile de se prononcer quant à une éventuelle amélioration de la productivité du travail tant que le projet n'est pas véritablement effectif. Pour la chaîne éditoriale le gain de temps sera sans doute nul. Par contre pour le traitement documentaire, c'est-à-dire l'extraction des notices bibliographiques, il y aura un gain de temps potentiel. Celui-ci sera d'autant plus important que le nombre de communications dans l'acte de colloque sera élevé. En tout cas, même si SGML</p>

	<p>n'améliore pas la productivité pour la chaîne éditoriale, nous pouvons nous attendre à une régularité de la qualité.</p>
<p>Améliorer la diffusion de la littérature grise du Cemagref en multipliant les supports de diffusion (papier, CD-Rom, en ligne). La littérature grise, au Cemagref, représente 60% de la production de textes. Cette amélioration passe notamment par une diffusion en ligne.</p>	<p>Oui, SGML peut améliorer la diffusion de la littérature grise. En effet, SGML permet de diffuser rapidement et facilement un même document sur plusieurs supports. Avec un seul et même fichier, on peut avoir diverses applications (impression, diffusion en ligne sur Internet). La possibilité de diffuser sur plusieurs supports, presque simultanément, permet d'accroître la diffusion de la littérature grise et donc de mieux la valoriser.</p>
<p>A travers une norme indépendante des moyens matériels, assurer un rôle de mémoire et de pérennité de l'information.</p>	<p>Oui, comme nous l'avons vu dans la deuxième partie, SGML est une norme indépendante des moyens matériels et logiciels. Nous pouvons donc penser que c'est un langage qui permet d'assurer un rôle de mémoire au Cemagref et de garder une pérennité de l'information.</p>
<p>Etudier l'acceptabilité de SGML pour les auteurs des communications. Ceux-ci devront-ils avoir des compétences spécifiques ?</p>	<p>Non, SGML n'est pas un langage qui peut être utilisé par des auteurs. Sa manipulation est trop complexe. Cependant comme nous l'avons vu précédemment on peut pallier à ce problème par le biais de feuille de style Word.</p>

Conclusion

Pendant ce stage, nous avons pu constater que le langage SGML n'est pas réservé qu'à des gros projets mais peut être utilisé pour des applications plus restreintes comme c'est le cas au Cemagref.

Pour la réussite de ce projet, la finalisation c'est à dire l'intégration dans la chaîne éditoriale devra se faire en complète concertation avec les personnes concernées (responsable d'édition, documentalistes, auteurs).

Il faut compter au moins six mois avant que l'application soit effective. Ce délai permettra également de mieux se rendre compte du coût effectif et du temps que cela nécessite avant d'étendre à d'autres types d'ouvrage l'utilisation de SGML.

De plus, dans l'avenir, on peut envisager d'autres applications telle que l'édition de CD-Rom à partir des documents SGML.

Bibliographie

EDITION ELECTRONIQUE

COMBEROUSSE, Martine. (1995). *Les nouvelles technologies au service de la littérature grise*. Bulletin Bibliographique de France, tome 40, p.51-53.

DUPOIRIER, Gérard. (1994). *Technologie de la GED : l'édition électronique*. Paris : HERMES. 228 p.

ISBN 2-86601-432-4

INSTITUT DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE. (1994).
Coordonné par Jean-Claude Le Moal et Bernard Hidoine. *Le traitement électronique du document : cours INRIA, 3-7 octobre 1994, Aix-en-Provence*. Paris : ADBS éditions. 287 p.

ISBN 2-901046-76-2 - ISSN 1160-2376

LUPOVICI, Catherine. (1995). *Le périodique électronique*. Bulletin Bibliographique de France, tome 40, p.32-37.

MAHAUT, Françoise, DALBIN Sylvie. (1992). *La gestion électronique de documents et la documentation : dossier documentaire*. Paris : ADBS éditions. 138 p. (Coll. : recherches et documents)

ISBN 2-901046-46-0

SGML

AFNOR, AIS/BERGER-LEVRAULT. (1995). SGML : Langage, stratégie et systèmes d'information. Stage du 30 au 31 mars 1995, Paris.

ARBORText. (1996). Getting start with SGML, [en ligne]. Adresse URL : <http://www.sgmlopen.org/sgml/docs/library/gestart.html>

COVER, Robin. (1996). Public SGML Software, [en ligne]. Adresse URL : <http://www.sil.org/sgml/publicSW.html>

COVER Robin, NAGGUM ERIK. (1994). Frequently asked question v1.7 : SGML, [en ligne]. Adresse URL : <ftp://info.ex.ac.uk/pub/SGML/faq/FAQ.txt>

DANISH STANDARDS ASSOCIATION. (1991). SGML-ODA : Présentation des concepts et comparaison fonctionnelle. Paris : Afnor. 87 p.

ISBN 2-12-488011-X

Rohrbach, Christian. (1996). SGML : Standardized General Markup Language, [en ligne]. Adresse URL : <http://www.iware.ch/SGML.htm>

SANDOVAL, Victor. (1994). SGML : un outil pour la gestion électronique de documents. Paris : Hermès. 174p. (coll. : Techniques de l'information)

ISBN 2-86601-440-5

VAN HERWIJNEN, Eric. (1995). SGML pratique : le langage utilisé par HTML, CALS, HyTime, DocBook. Paris : International Thomson Publishing France. 330 p.

ISBN 2-84180-009-1

NORMES

NF EN 28879. (1990). Traitement de l'information - Systèmes Bureautiques - Langage normalisé de balisage généralisé (SGML).

ISSN 0335-3931

ISO 12083. (1994). Information et documentation - Préparation et balisage de manuscrits électronique : Information and documentation - Electronic manuscript preparation and markup.

Liste de ressources SGML sur Internet

Les news

⇨ comp.text.sgml

C'est le plus important groupe de news sur SGML.

Les listes de diffusion

⇨ TEI-L

Liste de diffusion du Text Encoding Initiative, organisation internationale pour établir les principes pour l'encodage et l'échange des textes électroniques en utilisant SGML.

Envoyer un message à : LISTSERV@UICMV.CC.UIC.EDU

Texte du message : <sub TEI-L votre nom>

LES SERVEURS FTP

⇨ [ftp.sgm11.ex.ac.uk](ftp://ftp.sgm11.ex.ac.uk)

Login : anonymous

Password : votre E-mail

Informations sur SGML : les parseurs, bibliographie ; informations sur le projet SGML et sur le groupe d'utilisateurs de SGML du Royaume-Unis.

⇨ [ftp.ifi.uio.no](ftp://ftp.ifi.uio.no)

Login : anonymous

Password : votre E-mail

Service similaire à celui cité ci-dessus avec en plus des informations et les archives des messages postés sur comp.text.sgml

LES SERVEURS WEB

⇨ <http://www.sil.org/sgml/The SGML Web page>

Serveur le plus complet (faq, liste de logiciels, liste de publications, adresses des groupes d'utilisateurs...)

⇨ <http://www.sgmlopen.org>

Serveur du consortium international de SGML

⇨ <http://www.falch.no/~pepper/sgmltool>

Liste de logiciels SGML commerciaux et typologie

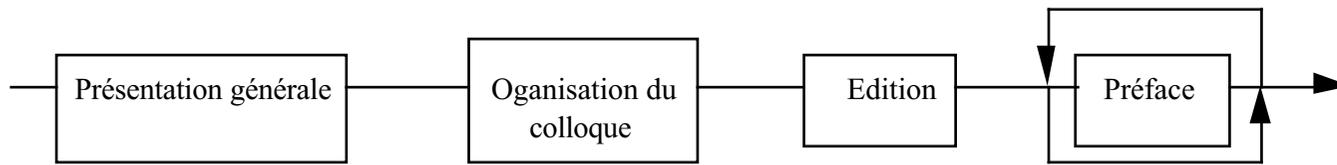
Annexes

Liste des annexes :

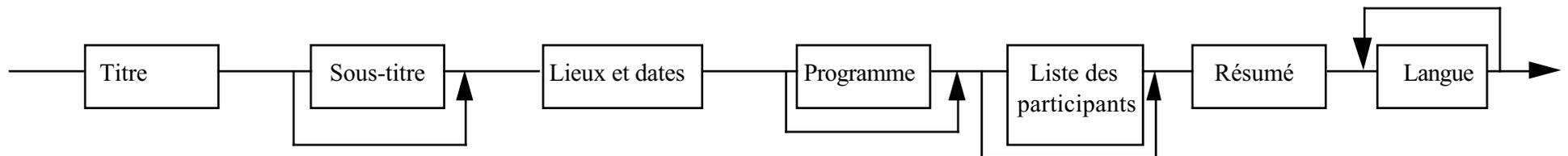
- ⇒ Diagrammes de structure de l'acte de colloque
- ⇒ Comparaison de la DTD Congrès avec la DTD Griseli et la notice documentaire
- ⇒ DTD Congrès

Diagrammes de structure de l'acte de colloque

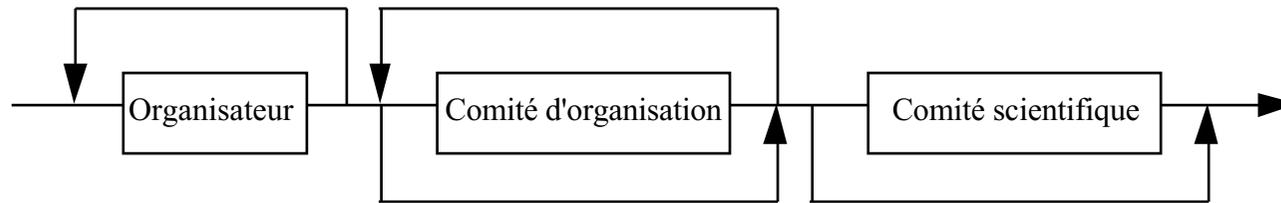
Colloque :



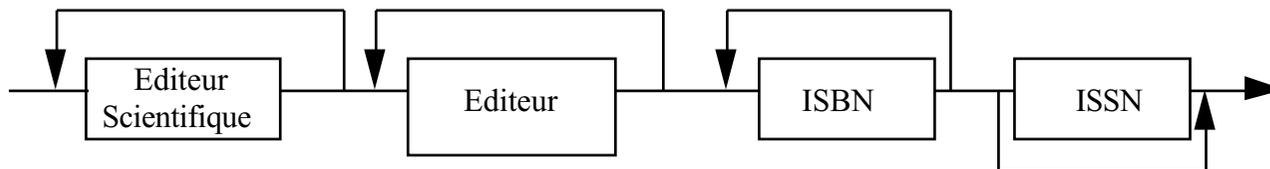
Présentation générale du colloque :



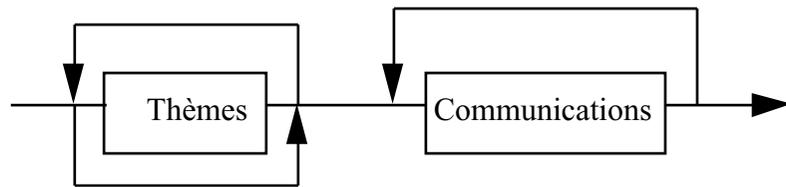
Organisation du colloque :



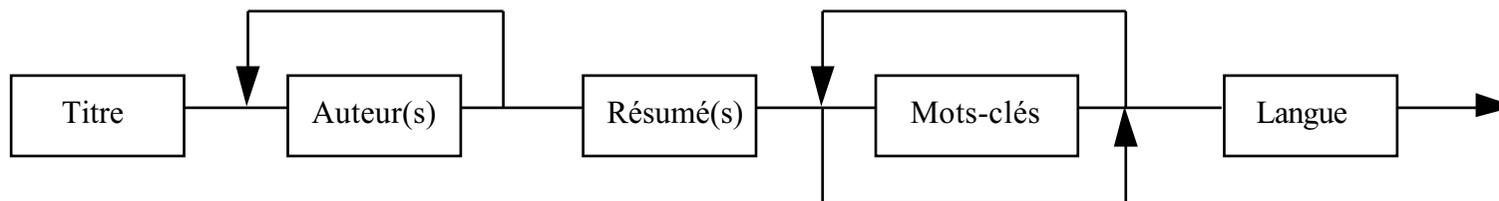
Edition du colloque :



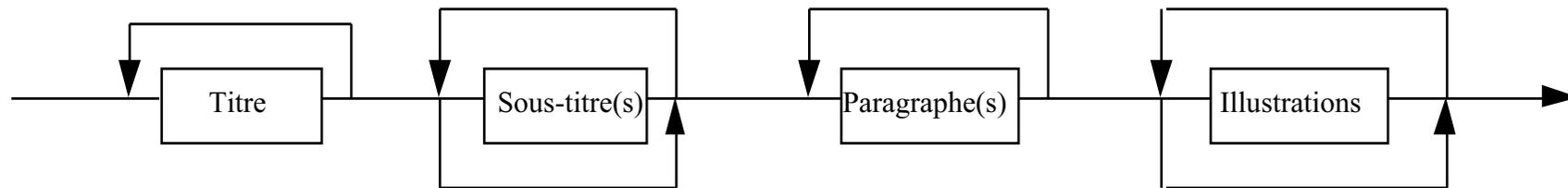
Session :



En-tête de la communication :



Corps de la communication :



Comparaison de la DTD Congrès avec la DTD Griseli et la notice documentaire du Cemagref

Eléments de la DTD actes de colloque	Élément de la DTD Griseli pour une notice acte de colloque	Élément de la DTD Griseli pour une notice communication	Notice documentaire du Cemagref pour une notice acte de colloque	Notice documentaire du Cemagref pour une notice communication
FRONT	1. FRONT	2. SUBFRONT	×	×
Titlegrp	1.1 Titlegrp (F)	2.1 Subtitlegrp (O)	×	×
Title	1.1.4 Title (O)	2.1.1 Title (O)	Titref (O), TitreA(O), TitreO (O)	×
Subtitle	1.1.5 Subtitle(F)	2.1.2 Subtitle (F)	×	×
Authgrp	1.2 Authgrp (F)	2.2 Authgrp (O)	×	×
Author	1.2.1 Author (F)	2.2.1 Author (O)	Auteur (O)	×
Corpauth	1.2.2 Corpauth (F)	2.2.2 Corpauth (F)	AuteurC (O)	×

Aff	1.2.3 Aff (F)	2.2.3 Aff (F)	AuteurC (O)	×
Pubfront	1.4 Pubfront (O)	1.4 Subpub (O)	×	×
Date	1.4.3 Date (O)	1.4.3 Date (O)	Date (O)	Date (O)
Pubname	1.4.4 Pubname (F)	1.4.4 Pubname (F)	Editeur (O, Nom)	Editeur (O), Nom
Location	1.4.5 location (O)	1.4.5 Location (O)	Editeur (O, Ville)	Editeur (O), ville
Confgrp	1.4.6 Confgrp (O)	1.4.6 Confgrp (O)	×	×
NO	1.4.6.1 NO (F)	1.4.6.1 NO (F)	Congrès (O) 1	Congrès (O) 1
Confname	1.4.6.2 Confname (O)	1.4.6.2 Confname (O)	Congrès (O) 2	Congrès (O) 2
Date	1.4.6.3 Date (O)	1.4.6.3 Date (O)	Congrès (O) 4, DateCGR (O, années)	Congrès (O) 4, DateCGR (O, années)
Location	1.4.6.4 Location (O)	1.4.6.4 Location (O)	Congrès (O) 3, Pays (O)	Congrès (O) 3, Pays (O)
Sponsor	1.4.6.5 Sponsor (F)	1.4.6.5 Sponsor (F)	×	×
ISBN	1.4.7 ISBN (F)	1.4.7 ISBN (F)	ISN (O)	ISN (O)
ISSN	1.4.8 ISSN (F)	1.4.8 ISSN (F)	ISN (O)	ISN (O)
Extent (nbre page)	1.4.12 Extent (O)	1.4.12 Extent (O)	Pages (O)	×

Abstract	4. Abstract (O)	×	ResumeF, ResumeA	×
Language	0.2.3 Language (0. Headrec, 0.2 Doc)	0.2.3 Language (0. Headrec, 0.2 Doc)	Langue	×
Groupe de travail	×	×	×	×
Intitulé	×	×	×	×
Préface	×	×	×	×
BODY	×	×	×	×
Thème	×	×	×	×
Sous-thème	×	×	×	×
Communication	×	×	×	×
<u>Comfront</u>	×	1. Front (O)	×	×
Titlegrp	×	1.1 Titlegrp (F)	×	×
Title	×	1.1.4 Title (O)	×	Titref (O), TitreA(O), TitreO (O)
Subtitle	×	1.1.5 Subtitle (F)	×	×

Authgrp	×	1.2 Authgrp (F)	×	×
Author	×	1.2.1 Author (F)	×	Auteur (O)
Corpauth	×	1.2.2 Corpauth (F)	×	AuteurC (O)
Aff	×	1.2.3 AFF (F)	×	AuteurC (O)
Abstract	×	4. Abstract	×	ResumeF, ResumeA
Keyword	×	×	×	×
Language	×	×	×	Langue
<u>Combody</u>	×	×	×	×
Introduction	×	×	×	×
Part	×	×	×	×
Chapter	×	×	×	×
Section	×	×	×	×
Subsect 1	×	×	×	×
Combiblist	×	×	×	×
Annexe	×	×	×	×

DTD Congrès

```
<!-- ***** -->
<!-- PUBLIC DOCUMENT TYPE DEFINITION SUBSET -->
<!-- ***** -->
<!DOCTYPE CONGRES
[
<!--
<!ENTITY % ISOlat1 PUBLIC "ISO 8879:1986//ENTITIES
Added Latin 1//EN" >
<!ENTITY % ISOlat2 PUBLIC "ISO 8879:1986//ENTITIES
Added Latin 2//EN" >
<!ENTITY % ISOamso PUBLIC "ISO 8879:1986//ENTITIES
Added Math Symbols: Ordinary//EN" >
<!ENTITY % ISOgrk1 PUBLIC "ISO 8879:1986//ENTITIES
Greek letters//EN" >
<!ENTITY % ISOgrk3 PUBLIC "ISO 8879:1986//ENTITIES
Greek letters//EN" >
<!ENTITY % ISOnum PUBLIC "ISO 8879:1986//ENTITIES
Numeric and special graphic//EN" >
<!ENTITY % ISOpub PUBLIC "ISO 8879:1986//ENTITIES
Publishing//EN" >
<!ENTITY % ISODia PUBLIC "ISO 8879:1986//ENTITIES
Diacritical Marks//EN" >
<!ENTITY % ISOTECH PUBLIC "ISO 8879:1986//ENTITIES
General technical//EN" >
-->
<!ENTITY % ISOlat1 SYSTEM "ISOLAT1" >
<!ENTITY % ISOlat2 SYSTEM "ISOLAT2" >
<!ENTITY % ISOamso SYSTEM "ISOAMSO" >
```

```
<!ENTITY % ISOgrk1 SYSTEM "ISOGRK1" >
<!ENTITY % ISOgrk3 SYSTEM "ISOGRK3" >
<!ENTITY % ISONum SYSTEM "ISONUM" >
<!ENTITY % ISOpub SYSTEM "ISOPUB" >
<!ENTITY % ISODia SYSTEM "ISODIA" >
<!ENTITY % ISOTech SYSTEM "ISOTECH" >
%ISolat1;
%ISolat2;
%ISOamso;
%ISOgrk1;
%ISOgrk3;
%ISONum;
%ISOpub;
%ISODia;
%ISOTech;
<!ENTITY % doctype "congres" -- document type generic identifier -->

<!-- ===== -->
<!-- ELEMENT DE BASE DU DOCUMENT -->
<!-- ===== -->
<!ENTITY % s.zz "(p)" -- Modele de paragraphe-->

<!-- ***** -->
<!-- MODELES -->
<!-- ***** -->
<!ENTITY % m.addr "(street?&city?&state?&country&postcode?&san?&email*&postbox?&phone*&fax*)"--
Modele d'element d'adresse -->
<!ENTITY % m.fig "EMPTY"-- Figure-->
<!ENTITY % m.name "(fname?&surname&degree*&school*&%m.addr;?&role*&aff?&qualif?)" -- Modele de
nom-->
<!ENTITY % m.org "(orgname&orgdiv*&%m.addr&role*)" -- Modele nom d'organisme-->
<!ENTITY % m.sec "(title?,(%s.zz;)*,(section)*)" -- Modele de section-->
```

```
<!-- ***** -->
<!-- DEFINITIONS DES ATTRIBUTS -->
<!-- ***** -->
<!ENTITY % a.id      "id id #IMPLIED" -- definition de l'attribut id -->
<!ENTITY % au.rid    "rid idrefs #IMPLIED" -- pour referer a un identifiant d'affiliation -->
<!ENTITY % iso639-2  -- codes langues --system "langue.txt" >
<!ENTITY % iso3166   system "pays.txt" -- codes pays -- >
<!ENTITY % a.types   "(latin|cyrillic|greek|hebrew) latin" -- alphabet -- >
<!ENTITY % d.types   -- types de date --
-- 6 : aaaa --
-- 7 : aaaamm --
-- 8 : aaaammjj --
"(6|7|8) #REQUIRED" >
<!ENTITY % a.grtypes "(sient|organ|autre)" -- Modele de type de groupe de travail-- >
<!ENTITY % m.toc     "empty" -- Modele de tables des matieres-->

<!-- ===== -->
<!-- STRUCTURE DU DOCUMENT -->
<!-- ===== -->
<!ELEMENT %doctype; - - (front,body,annexe*) >

<!-- ===== -->
<!-- EN-TETE ACTE DE COLLOQUE -->
<!-- ===== -->
<!ELEMENT front      - - (titlegrp,authgrp,pubfront,abstract+,language+,grptravail?,programme?, toc?) --
En-tete de l'acte de colloque-->

<!--+++++++ -->
<!-- GROUPE TITRE (PARTIE MONOGRAPHIQUE) -->
<!-- ++++++ -->
<!ELEMENT titlegrp  - - (title,subtitle)* -- Groupe titre-- >
<!ELEMENT title     - - (#pcdata) -- Titre-->
<!ELEMENT subtitle  - - (#pcdata) -- Sous-titre-- >
<!-- ++++++ -->
```

```
<!-- GROUPE AUTEUR -->
<!--+++++++>
<!ELEMENT authgrp -- (author,corpauth,aff)* -- Groupe Auteur-- >
<!ELEMENT author -- (%m.name;) -- Auteur-- >
<!ELEMENT corpauth -- (%m.org;) -- Collectivite auteur-->
<!ELEMENT aff -- (%m.org;) --Affiliation-- >

<!-- ++++++ -->
<!-- PUBLICATION DE LA MONOGRAPHIE -->
<!-- ++++++ -->
<!ELEMENT pubfront -- (date,pubname+,location,confgrp,isbn*,issn?,extent) -- Publication-->
<!ELEMENT date -- (#pcdata) -- date d'edition-->
<!ELEMENT pubname -- (%m.org;) -- Nom de l'editeur-->
<!ELEMENT location -- (%m.addr;) -- Adresse editeur-->
<!ELEMENT confgrp -- (no?,confname,date,location,sponsor*) --Groupe congres-->
<!ELEMENT no -- (#pcdata) -- Numero du congres-->
<!ELEMENT confname -- (#pcdata) -- Nom du congres-->
<!ELEMENT sponsor -- (%m.org)-- organisateur-->
<!ELEMENT isbn -- (#pcdata) --Isbn-- >
<!ELEMENT issn -- (#pcdata) --Issn-- >
<!ELEMENT extent -- (#pcdata) -- Nombre de pages-->

<!-- ++++++ -->
<!-- RESUME -->
<!-- ++++++ -->
<!ELEMENT abstract -- (p+) -- Resume du Congres-- >

<!-- ++++++ -->
<!-- LANGUE -->
<!-- ++++++ -->
<!ELEMENT language -- (#pcdata) -- Langue de l'acte de colloque-- >

<!-- ++++++ -->
<!-- ELEMENTS ACTE DE COLLOQUE -->
```

```
<!-- ++++++ -->
<ELEMENT grtravail -- (intitule,authgrp)-- Groupe de travail-->
<ELEMENT intitule -- (#pcdata)-- Intitule-->
<ELEMENT programme -- (title,p)*-- Programme du congres-->

<!-- ++++++ -->
<!-- TABLE DES MATIERES -->
<!-- ++++++ -->
<ELEMENT toc -- %m.toc; -- Table des matieres -->

<!-- ===== -->
<!-- CORPS DE L'ACTE DE COLLOQUE -->
<!-- ===== -->
<ELEMENT body -- (theme+) -- Corps du congres-->
<ELEMENT theme -- (no?, title?,(soustheme|communication)*)-- Theme-->
<ELEMENT soustheme -- (title?, (communication)*)-- Sous-theme*-->
<ELEMENT communication -- (comfront,combody,combiblist?)-- Communication-->
<ELEMENT comfront -- (titlegrp,authgrp,abstract+,language,keyword*)-- Preliminaire de la
communication-->
<ELEMENT keyword -- (#pcdata)-- Mot-cle-->
<ELEMENT combody -- (part+)-- Corps de la communication-->
<ELEMENT part -- (no?,title?,(%s.zz;)*,chapter+) -- Partie-->
<ELEMENT chapter -- (no?,%m.sec;) -- Chapitre-->
<ELEMENT section -- (no?,title?,(%s.zz;)*,subsect1*) --Section-->
<ELEMENT subsect1 -- (no?,title?,(%s.zz;)*,subsect2*)-- Sous-section1-->
<ELEMENT subsect2 -- (no?,title?,(%s.zz;)*,subsect3*)-- Sous-section2-->
<ELEMENT subsect3 -- (no?,title?,(%s.zz;)*,subsect4*)-- Sous-section3-->
<ELEMENT subsect4 -- (no?,title?,(%s.zz;)*,subsect5*) -- Sous-section4-->
<ELEMENT subsect5 -- (no?,title?,(%s.zz;)*,subsect6*)-- Sous-section5-->
<ELEMENT subsect6 -- (no?,title?,(%s.zz;)* -- Sous-section6-->
<ELEMENT p -- (#pcdata) -- Paragraphe-->
<ELEMENT combiblist -- (titlegrp,authgrp,date?,pubname?,location?,confgrp?,
isbn?,issn?,extent?)* -- Bibliographie-->
```

```
<!-- ===== -->
<!-- ANNEXE -->
<!-- ===== -->
<!ELEMENT annexe -- (no?,title*,p*) --Annexes-->

<!-- ***** -->
<!-- ELEMENTS DIVERS -->
<!-- ***** -->
<!ELEMENT orgname -- (#pcdata) -- Nom de l'organisme-- >
<!ELEMENT orgdiv -- (#pcdata) -- Nom de la division de l'organisme-->
<!ELEMENT fname -- (#pcdata) --Prenom-->
<!ELEMENT surname -- (#pcdata) -- Nom-- >
<!ELEMENT role -- (#pcdata) --role-- >
<!ELEMENT qualif -- (#pcdata) -- qualificatif du nom-->
<!ELEMENT street -- (#pcdata) -- rue-- >
<!ELEMENT city -- (#pcdata) -- ville-->
<!ELEMENT state -- (#pcdata) --Etat-- >
<!ELEMENT country -- (#pcdata) -- Pays-- >
<!ELEMENT postcode -- (#pcdata) --Code postal-->
<!ELEMENT san -- (#pcdata) -- Numero d'adresse standard-- >
<!ELEMENT email -- (#pcdata) -- email-- >
<!ELEMENT mailbox -- (#pcdata) -- Boite postale-->
<!ELEMENT phone -- (#pcdata) --Telephone-- >
<!ELEMENT fax -- (#pcdata) --fax-- >
<!ELEMENT degree -- (#pcdata) --degre-->
<!ELEMENT school -- (%m.org;) --Ecole-->

<!-- ===== -->
<!-- DEFINITION DES ATTRIBUTS -->
<!-- ===== -->
<!ATTLIST date datetype %d.types>
<!ATTLIST intitule types %a.grtypes #required >
<!ATTLIST grptravail grptype (comite|partic) #required >
<!ATTLIST country cnycode %iso3166 #required >
```

```
<!ATTLIST author                %au.rid; >
<!ATTLIST aff                    %a.id; >
<!ATTLIST title lngcode %iso639-2 #required
titype (1|2) #required
alphabet %a.types >
<!ATTLIST subtitle lngcode %iso639-2 #required
titype (1|2) #required
alphabet %a.types>
<!ATTLIST language lngcode %iso639-2 #required>
<!ATTLIST confname
lngcode %iso639-2 #required
alphabet %a.types >
<!ATTLIST abstract
lngcode %iso639-2 #required>
<!ATTLIST keyword
lngcode %iso639-2 #required>
]>
```