

PRODUIRE, DIFFUSER, PARTAGER DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE

*SOUS QUELLES FORMES ? SUR QUELS SUPPORTS ?
AVEC QUEL OUTILS ?*

Depuis 1960, l'imprimerie Louis-Jean, située à Gap dans les Hautes-Alpes, a cherché les meilleures solutions pour répondre aux besoins de communication de l'information émise par le milieu scientifique.

L'évolution des techniques de composition ainsi que celles des transmissions télématiques ont été tellement rapides que l'on oublie trop souvent qu'il y a à peine 20 ans la micro-informatique faisait seulement de timides incursions dans le monde scientifique. Quant au monde des imprimeurs, il criait au scandale si l'on osait s'aventurer à composer des textes avec de tels outils synonymes à l'époque d'anti-professionnalisme.

Apparition de nouveaux supports

Quelques imprimeurs dont Louis-Jean testèrent très rapidement – au risque de déplaire – ces nouveaux outils qui, à l'instar des systèmes de composition traditionnels et « fermés », avaient pour caractéristique principale d'être ouverts sur le monde extérieur.

L'information pouvait enfin être produite dans le bureau ou le laboratoire d'un scientifique, puis transmise numériquement au professionnel qu'est l'imprimeur, afin de l'impri-

mer sur du papier. Le premier intérêt directement perçu par les éditeurs fut donc tout simplement l'économie générée par la récupération des textes « auteurs » sous forme électronique.

Progressivement, le monde des scientifiques et celui des imprimeurs circonscrivent alors une zone, où chacun pouvait échanger avec plus ou moins de bonheur de l'information numérique.

Et, durant ces dernières années, la problématique d'un éditeur en matière de fabrication était de déterminer jusqu'où l'auteur ou le secrétaire d'édition devait aller et où l'imprimeur devait reprendre les données, ceci dans un souci d'efficacité et d'économie optimales.

Dans un second temps, ce ou ces nouveaux supports ont laissé entrevoir de nouveaux modes d'accès à l'information. Les pratiquants, qu'ils soient auteurs ou opérateurs, découvraient les plus-values offertes par un traitement numérique. Le papier, dans la culture de tous, n'était plus l'unique support possible. Les produits numériques pouvaient devenir une fin en soi, une autre façon d'appréhender l'information. Et la nature de l'information scientifique elle-même convenait parfaitement à cette évolution.

FRANÇOISE FAUGERAS
Imprimerie Louis-Jean

fonction du lectorat visé, et en publiant tout ou partie suivant des critères donnés ;

– partager l'information à distance en utilisant une logique de réseau.

A chaque fonctionnalité correspondent des outils et des normes, car dès lors que l'on décide de partager, il est utile de s'y référer.

L'archivage

Si l'archivage « papier » fait l'unanimité et ne pose pas de gros problème d'accès ou de pérennité, on ne peut en dire autant des formats de fichiers numériques et des supports utilisés. Combien d'imprimeurs et d'éditeurs ont dû recomposer des textes numériques, car le support était inutilisable, ou parce que la dépense d'énergie pour les passer dans un format actuel dépassait les coûts de recomposition.

Pour pérenniser l'information, deux pratiques sont actuellement utilisées.

La première assez pragmatique est de développer en permanence les outils nécessaires au passage d'un format propriétaire à un autre et d'un support à un autre. Bien évidemment, plus les « composants » de l'information sont variés, plus l'opération est acrobatique et périlleuse à moyen et long terme.

La seconde est de se référer à une norme ISO (International Organization for Standardization) et de produire son fonds documentaire sous cette norme. La référence à une norme pose *a priori* à l'éditeur un problème de coût. Normaliser une information induit des économies à moyen et long terme, soit parce que la mise en forme sera facilitée ou automatisable, soit parce que la récupération sera immédiate, mais c'est une procédure plus coûteuse au départ. En effet, la norme utilisée, le SGML ou Standard Generalized Markup Language, nécessite un gros travail de mise en architecture.

Créée en octobre 1986, pour la représentation des documents, cette norme a pour idée directrice de séparer la ou les formes du fond. Un document SGML est un contenu associé à une structure logique ou DTD

(Document Type Definition). Concevoir une DTD prend du temps et nécessite une compétence particulière. L'éditeur ou l'imprimeur ne fera cet investissement que si le bénéfice retiré est supérieur au surcoût engendré.

Parallèlement, l'éditeur ne détermine pas forcément complètement sa politique éditoriale et sa stratégie commerciale en fonction des nouveaux outils mis à sa disposition. Il reste encore beaucoup de travail pédagogique et d'échanges à faire entre les éditeurs définissant leurs projets éditoriaux et les prestataires de services maîtrisant et développant les outils, afin d'exploiter au mieux les possibilités de diffusion multisupports.

Nous vivons donc en ce moment une période d'équilibrage où se testent les nouveaux outils aussi bien que les nouveaux modes de consommation. Période cruciale durant laquelle le partenariat entre éditeur et imprimeur doit être plus fort que jamais. Chacun a besoin de l'autre pour évoluer du bon côté et trouver les bonnes solutions.

Pratiquement, un éditeur ou un imprimeur, pour produire de l'information normée SGML, utilisera dans un premier temps un éditeur SGML, qui permet non seulement d'intégrer les balises de structure nécessaires, mais de les relier de façon logique et, enfin, d'en valider la conformité par rapport au modèle construit.

Un éditeur SGML peut être plus ou moins convivial et facile d'apprentissage. Il peut accepter ou non de travailler avec des documents non valides sans bloquer l'opérateur. Le choix de cet éditeur est donc particulièrement important suivant les compétences des opérateurs, les assistances techniques possibles et la productivité souhaitée (*cf.* encadré ci-contre).

Un éditeur doit en outre le plus souvent récupérer une collection antérieure et la passer au format SGML. Trois techniques sont possibles : nouvelle saisie des textes, lecture optique ou récupération par transcodage et intégration du balisage SGML en exploitant au mieux les repères typographiques pour automatiser cette tâche et/ou en utilisant des logiciels

spécifiques de conversion (Dynatag, Omnimark, Balise)

Une fois que l'instance SGML est produite, le problème de l'archivage est réglé, mais l'information n'est pas accessible au commun des lecteurs, puisqu'elle est sans formes. Il s'agit alors de la rendre disponible pour l'utilisateur *lambda*.

L'accès à l'information

Si l'éditeur a choisi une méthode de production de l'information apte à s'exporter sur tous supports, il est bien entendu qu'à partir de sa base documentaire SGML, des outils ont dû être développés pour permettre au lecteur moyen d'accéder à l'information dans des conditions optimales suivant le support qu'il aura choisi.

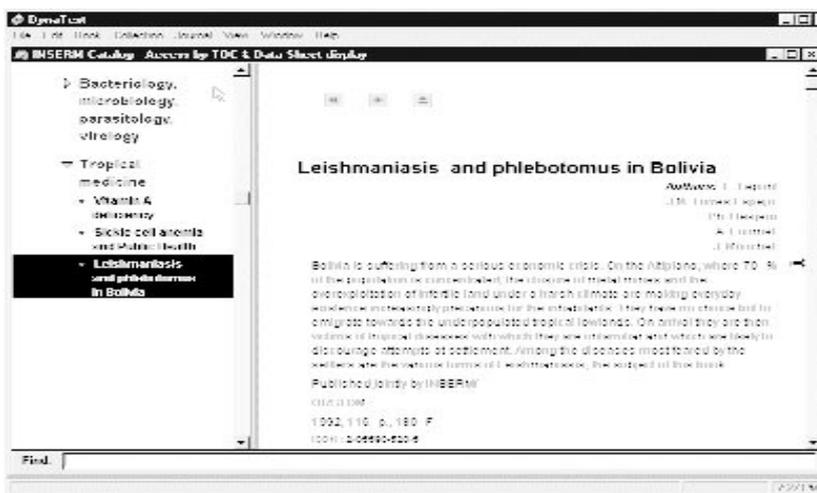
S'il s'agit du papier, l'éditeur aura dû, pour rendre l'information « lisible », utiliser un logiciel de mise en pages capable d'intégrer du SGML et d'en récupérer toutes les informations. Framemaker+SGML est un de ces outils. De plus en plus utilisé pour traiter du texte structuré, cet outil a l'avantage d'être convivial et son apprentissage peut être progressif.

Framemaker+SGML possède deux « briques », la première (Framemaker) permettant la composition et la mise en pages, la seconde (+SGML) permettant d'accoler à l'information une structure logique (*cf.* planche 1). S'il s'agit d'écran, l'éditeur devra faire appel à un visualiseur. Pour une consultation sur CD-ROM, le visualiseur possédera des qualités spécifiques à ce support ; pour une consultation *via* Internet, d'autres qualités seront exi-

Éditeurs SGML les plus répandus

Adept Editor et Adept Publisher (Arbortext)
Author/Editor (Softquad)
Framemaker+SGML
Near&Far Author (Microstar)
SGML Tag Wizard
Grif SGML Editor
Interleaf 6 SGML

Planche 2
Livre Dynatext



gées. Dynatext/Dynaweb sont deux outils de visualisation utilisés à l'imprimerie Louis-Jean (cf planche ...).

Les visualiseurs les plus répandus

- Dynatext/Dynaweb (Ebt/Inso)
- GTI Publisher (Jouve)
- Folioviews (Opentext)
- Insight into information (Enigma)

Les recherches multicritères

Tout l'intérêt d'avoir structuré l'information est de pouvoir repérer à tout moment des éléments définis dans la structure. Que ce soit pour un export « papier » ou électronique, dès lors que la structure a été convenablement définie, il est possible d'éditer des listes simples d'éléments de même nature (par exemple : liste d'auteurs, de mots-clés, de figures, de tableaux etc.) ou des listes plus complexes produites par le croisement de deux éléments (liste des auteurs ayant écrit un article sur tel mot-clé, liste des tableaux liés à tel thème, etc.).

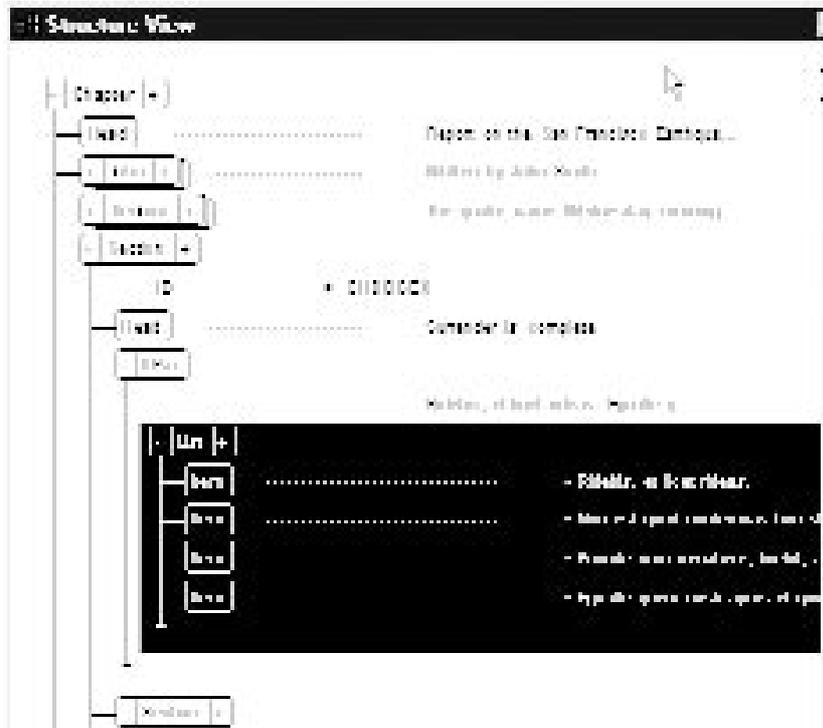
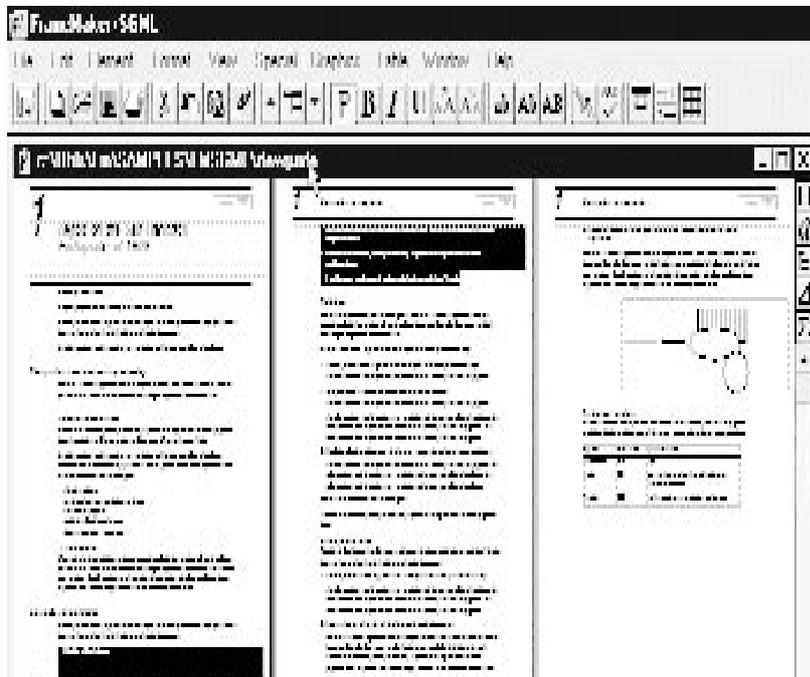
Sur un support numérique, la problématique de diffusion est différente, puisque, à tout moment, le lecteur peut, si le visualiseur le permet, construire une requête basée sur la structure de la base documentaire SGML.

Ici naît une des plus-values les plus appréciables de la production électronique : la liberté donnée au lecteur de choisir son mode d'accès à l'information.

Dans un livre électronique de type Dynatext, le lecteur peut circuler dans de multiples tables de matières ou *via* le moteur de recherches (cf planche ...).

D'autre part, la navigation dans le livre électronique devient facile grâce à la création de liens prédéfinis ou même à celle de liens personnalisés et construits directement par le lecteur. Enfin, certains visualiseurs comme Dynatext offrent au lecteur la possibilité d'accéder au son, à l'image ani-

Planche 3
Framemaker+sgml, vues de la structure et de la mise en page associées



mée ou même au déclenchement d'applications extérieures, ce qui n'est pas du tout envisageable dans le cas de l'ouvrage sur papier.

L'exploitation d'une base documentaire sous toutes ses formes

Un des atouts les plus remarquables, et non encore exploité actuellement de façon vraiment optimale chez les éditeurs, est de penser la structure de sa base documentaire de telle manière que l'information puisse être exportée globalement ou partiellement pour un niveau déterminé de lecture.

Il est très aisé et rapide de publier un ensemble d'informations extraites d'une base et répondant à tel ou tel critère si l'information est structurée. Par exemple un éditeur scolaire peut, à partir d'une base d'exercices sur un thème donné, produire un ensemble « exercices seuls », un ensemble « exercices et corrigés », un ensemble « exercices et cours associés », etc.

En ce qui concerne la forme pure, et pour des besoins de stratégie commerciale, un éditeur pourra aussi publier la même information sous différents « visages », afin d'atteindre plusieurs cibles.

Le partage de l'information à distance

Si, jusqu'à présent, l'éditeur concevait sa production éditoriale pour une consommation sur un lieu géographique donné, la progression d'Internet lui permet dorénavant de

Glossaire

HTML (Hypertext Markup Language) : langage de description de pages Web directement issu du langage SGML. (HTML est une DTD SGML).

SGML (Standard Generalized Markup Language, ISO 8879 : 1986), langage normalisé de balisage généralisé. Permettant de décrire la structure d'un document, ce langage de représentation formalise le balisage et le rend indépendant des systèmes et des traitements.

XML (Extensible Markup Language), langage de balisage de documents, version simplifiée de SGML destinée aux applications Internet, mais plus performante que HTML.

DTD (Document Type Definition) : définition de type de document. Définition des règles de balisage pour une classe donnée de documents. Chaque document conforme à SGML doit comporter une DTD ou une référence à une DTD.

penser autrement. Les scientifiques utilisent le réseau depuis longtemps maintenant pour échanger des données non validées par leurs pairs. L'enjeu pour les éditeurs scientifiques est de déterminer leur spécificité sur le réseau. Un auteur peut librement échanger sur Internet. Un éditeur doit apporter une plus-value à l'information, soit parce qu'elle sera validée, repérée ou accessible plus facilement, soit parce que sa mise en forme sur le réseau sera plus attractive.

Les visualiseurs Web utilisent la norme HTML et utiliseront bientôt la norme XML (Extensible Markup Language). C'est une des raisons du succès grandissant de SGML que de répondre à tous les besoins (papier, CD-ROM, Web). HTML ou XML n'étant que des sous-produits de SGML, utiliser ce dernier comme plate-forme de production de l'information devient une opération de plus en plus rentable. Bon nombre d'éditeurs scientifiques ou techniques optent maintenant pour cette norme en raison de cette plurifonctionnalité.

Certains outils de diffusion sur le Web exploitent les avantages de la structure SGML des documents. Ainsi un outil comme Dynaweb (Ebt/Inso) permet de produire des sites Web dynamiques, où le lecteur va pouvoir en temps réel rapatrier l'information qu'il cherche suivant des critères précis (cf. fig. 5 : site Web dynaweb).

On le voit, l'évolution en matière de traitement de l'information est rapide. L'imprimeur, soucieux de répondre aux besoins de ses éditeurs, cherche en permanence l'association des meilleurs outils pour produire et diffuser en faisant l'économie optimale.

Comment évoluer rapidement, efficacement, comment développer le maximum de créativité pour cette nouvelle donne ? C'est le défi que tout acteur de la chaîne de production éditoriale doit se fixer aujourd'hui. C'est autour de cet objectif qu'éditeurs et imprimeurs doivent rassembler leurs forces pour gagner ce pari.

Mars 1997