

D'Etebac à XML : organiser la cohabitation des langages

LES BESOINS DES CLIENTS DES banques sont fondamentalement différents suivant la nature même de ces derniers (particuliers, PME, grandes entreprises), suivant la nature des transactions

qu'ils utilisent et, bien entendu, suivant le marché dans lequel ils s'inscrivent, domestique ou international par exemple.

Pour satisfaire au mieux ces besoins, les communautés bancaires ont été amenées à proposer une offre très diversifiée dans pratiquement tous les domaines de leur activité, et par voie de conséquence à adapter leurs modes de communication avec leurs clients. Ceci n'a d'ailleurs pas été de la seule initiative des banques, les clients – notamment les grandes entreprises – constituant également dans ce domaine une force de proposition importante.

DE PLUS EN PLUS D'INFORMATIONS À UN COÛT DE MOINS EN MOINS ÉLEVÉ

Lorsque l'on parle de communication client-banque, plusieurs facteurs déterminant rentrent en ligne de compte :

- le moyen de communication (papier, téléphone, télématique, etc.) : ce poste est en progression constante du fait de l'évolution de l'environnement technique de plus en plus rapide dans ce domaine ;
- le contenu de la communication :

Les relations télématiques entre les banques et leurs clients notamment entreprises, sont passées du simple minitel à des langages plus évolués comme Etebac, Edifact et bientôt XML.

Les banques sont contraintes de gérer la coexistence de ces différentes syntaxes.

pour partie dépendant du moyen utilisé (capacité à transporter tel ou tel type d'information), ce point est majoritairement lié, dans sa définition même, à la satisfaction des besoins des deux parties en présence, en l'occurrence le client et sa banque. Il faut également noter sur ce plan que le contenu de l'information à acheminer dans le cadre d'une transaction financière sera vraisemblablement d'autant plus complexe que la transaction elle-même est complexe. Un autre facteur, à résonance aussi bien économique que technologique, réside dans la capacité des systèmes d'information à véhiculer de plus en plus rapidement, de plus en plus d'information, le tout à un coût de moins en moins élevé. Il en résulte en effet une acceptation, généralisée à l'ensemble des acteurs, à se transmettre des informations plus détaillées et donc plus riches ;

- le langage utilisé pour communiquer : ici intervient un élément majeur puisque la définition des deux premiers facteurs précités n'a de sens que si les acteurs ont convenu d'un langage commun pour communiquer entre eux. En

termes informatiques, c'est-à-dire la capacité à comprendre et traiter une information reçue par voie électronique, ce besoin vital se traduit – au moins schématiquement – par deux contraintes.

Il faut d'une part des définitions de données communes. A l'instar des mots composant une langue écrite et parlée, il convient que les données – équivalent des mots – transmises aient la même signification pour tous. On verra plus loin

“ A l'instar d'une langue écrite et parlée, il convient que les données transmises aient la même signification pour tous. ”

que le concept de donnée, initialement unitaire (donnée élémentaire), s'est élargi à des groupes de données composant un ensemble utilisable tel quel. Ce nouveau concept est maintenant fréquemment qualifié par le terme « objet » ou en anglais *core component*. A titre d'exemple simple et couramment utilisé, l'objet « adresse » pourra être composé d'un numéro de rue, d'une rue, d'une ville, d'un code postal, d'un pays, etc. Cette pré-



PIERRICK DAUMAIN

Chargé de mission affaires européennes

Comité français d'organisation et de normalisation bancaires (CFONB)

1. Edifact, l'esperanto du langage électronique

Créé en 1987 sur une initiative de l'ONU, l'Edifact (Echange de données informatisées pour faciliter l'administration, le commerce et les transports) était destiné à permettre la dématérialisation de documents-papier. Un travail considérable de définitions de données (plus de 10 000), de syntaxe et de règles d'utilisation a été fourni, concrètement

par la production de certains de messages utilisés dans presque tous les domaines, y compris la finance bien entendu.

A considérer un peu comme l'«esperanto» du langage électronique, l'EDIFACT présente, pour chacun de ses messages, du fait notamment qu'il couvre des besoins à l'échelle mondiale, une telle richesse de possibilités d'utilisation que sa mi-

se en œuvre concrète entre deux parties nécessite pratiquement toujours :

- l'adaptation du message à l'usage domestique, par exemple, en France le champ «identification du compte» sera renseigné par le RIB ;
- des accords bilatéraux préalables et un investissement non négligeable. Paradoxalement, c'est la richesse des possibilités

d'utilisation des messages qui constitue l'un des principaux obstacles au développement de l'EDIFACT dans le domaine financier. En effet, la part de l'information strictement liée au paiement dans une transaction commerciale est souvent infime par rapport à la documentation commerciale proprement dite (descriptif des articles, facture détaillée, etc.).

cision est importante dans la mesure où, comme on le verra plus loin, elle a un impact fort sur l'émergence de nouveaux langages et notamment d'XML (*eXtended Markup Language*) qui a pris une importance prépondérante via le développement d'internet.

Il faut d'autre part une syntaxe commune, c'est-à-dire un mode d'organisation et de présentation des données (la «grammaire» qui accompagne le dictionnaire) reconnue par les deux parties qui conversent entre elles. Ceci n'empêche bien évidemment pas la coexistence de plusieurs syntaxes dès lors que, soit celles-ci sont universellement connues, soit les deux parties ont préalablement convenu celle qu'ils vont utiliser. De la même manière, l'utilisation de telle ou telle syntaxe ne devrait pas constituer un obstacle à l'utilisation du même dictionnaire de données, celui-ci constituant, par sa nature de socle commun, un acquis essentiel.

Sur un plan plus général, afin d'optimiser la prise en compte de ces deux facteurs, et surtout de mieux les insérer dans un souci de cohérence globale entre les besoins des différents intervenants dans une transaction, la nécessité est apparue d'inclure le processus de

conception des messages dans une modélisation de la transaction beaucoup plus globale.

LA NÉCESSITÉ D'UNE MODÉLISATION PLUS GLOBALE DES TRANSACTIONS

Concrètement, la modélisation d'une transaction permet de répondre à deux préoccupations :

- se doter de la capacité à assurer l'intégrité de l'information de bout en bout : les besoins en informations de tous les intervenants tels qu'ils ressortent du modèle doivent pouvoir être identifiés et satisfaits ;
- fonctionner en STP (pour *Straight Through Processing*) c'est-à-dire la capacité à assurer cette intégrité de bout en bout de manière totalement automatisée en mettant en place les moyens appropriés.

Pour que ces deux résultats soient atteints, il faut donc que le processus d'élaboration des messages s'inscrive dans un modèle global, seul moyen d'identifier les besoins de tous les acteurs. Il faudra de plus que ce modèle prenne en compte les éventuelles contraintes techniques liées au mode de transport (capacité et contraintes de tel ou tel ré-

seau) et à la sécurité des échanges, facteurs particulièrement importants dans toute transaction impliquant une opération financière.

ÉTAT DES LIEUX «AVANT XML»

Dans les échanges de données informatisées (EDI), on a vu que l'utilisation de la plupart des syntaxes implique soit une normalisation pointue des données (et de leurs définitions) pour qu'elles soient reconnues et utilisables par l'ensemble des parties communicantes, soit l'établissement de relations bilatérales entre deux parties pour qu'aucune ambiguïté ne puisse exister quant à la signification à donner au contenu des messages échangés.

Le dialogue télématique client-banque en France s'est développé dans les années 80 sur la base de deux principaux courants :

- le minitel est utilisé indifféremment par les entreprises et les particuliers. L'absence d'utilisation de cet outil dans l'interbanquarité a fait qu'il n'a pas donné

“ Assurer l'intégrité de l'information de bout en bout de manière totalement automatisée. ”

Banque d'entreprise

lieu, dans ce cadre, à des travaux de normalisation ;

- Etebac est, de son côté, uniquement destiné aux entreprises et utilisable uniquement à partir de PC ou d'unité informatique. Ce type de relation n'a pu être établi qu'à l'issue d'un processus de normalisation fort, tant au niveau du mode de communication que des formats des messages utilisables.

Ces deux processus (minitel et Etebac) sont restés limités au marché domestique français, à quelques rares exceptions comme, par exemple, les clients à l'étranger de banques françaises transmettant des fichiers Etebac.

2. Exemples

- Le texte à l'écran suivant :

«Brochures CFONB

Répertoire des normes»

- Ecrit en HTML, étant entendu que

<H1> signifie titre 1

 et début et fin de ligne

→ <H1>Brochures CFONB</H1>

Répertoire de normes

- Ecrit en XML, étant entendu que des balises ont été définies comme, à titre d'exemple,

<titre> est le titre d'un livre, etc.

<Biblio=Brochures CFONB>

<Titre>Répertoire des normes</Titre>

</Biblio>

Au plan international, si le dialogue entre banques s'est considérablement développé grâce notamment et principalement à Swift, une seule norme restait internationalement reconnue pour le dialogue client-banque : l'Edifact (*encadré 1*). Dans les faits, compte tenu de la lourdeur des investissements nécessaires pour implémenter l'Edifact, d'une part, et de la très large diffusion par les fournisseurs de progiciels d'outils intégrant Etebac et les formats plats CFONB, d'autre part, l'utilisation de la norme Edifact dans les échanges client-banque n'a connu de développement qu'avec les

grandes entreprises à vocation internationale (et à leur demande le plus souvent).

LA NAISSANCE D'HTML

Sur la base de ces constats, compte tenu du développement du marché des ordinateurs individuels capables de communiquer avec l'extérieur chez les PME et les particuliers, un certain nombre d'offres, privatives bancaires ou de la part de SSII, ont vu le jour. En parallèle à la diffusion de ces progiciels, un processus de simplification et d'harmonisation des langages utilisés a été engagé par l'ensemble des acteurs intéressés, afin

notamment de réduire les coûts de développement de ces produits. Dans le même esprit, compte tenu de la diversité des demandes de la clientèle dans ce domaine, il est apparu nécessaire de construire des ensembles cohérents permettant la transmission, au choix, de texte, de son, d'image. Par ailleurs, le développement du web, par nature utilisé par de multiples acteurs ne se connaissant pas, a généré le besoin de définir des fichiers comportant, outre les données à présenter, leur mode de pré-

sentation sur écran. C'est de ce besoin qu'est né HTML (*Hyper-Text Markup Language*) que l'on peut considérer comme l'ancêtre d'XML.

Le principe générique d'HTML est de «baliser» chaque donnée, ou groupe de données, par des balises indiquant de quelle manière cette donnée doit être mise en page sur papier ou présentée à l'écran. Ces balises peuvent être considérées comme passives («non intelligentes») puisqu'elles n'apportent aucune indication quant à la signification de la donnée, et n'ont donc aucune valeur sémantique.

En systématisant le principe de l'utilisation de balises normalisées, HTML permet, de manière simple, de transmettre un texte et son mode de représentation pour chaque donnée ou groupe de données : un titre reste un titre, une ligne reste une ligne, un format en italique pour une portion de texte est maintenue, etc. Une carence de taille reste toutefois dans l'absence totale de contenu sémantique de la donnée, c'est-à-dire de la signification à lui attribuer. C'est pour pallier cette carence qu'à notamment été créé XML.

XML, UN SYSTÈME DE BALISES INTELLIGENTES

XML s'est donc construit sur la base de balises «intelligentes», c'est-à-dire permettant de formater le texte mais aussi d'explicitier le contenu du texte «balisé». Grâce à cette évolution, les balises prennent un sens et peuvent même être significatives d'une hiérarchie dans le texte. Ces balises doivent aussi respecter un certain nombre de règles, que l'on pourrait qualifier de «grammaire» XML dont, à titre d'exemple, les informations sont encadrées de balises «ouvrantes» et «fermantes», elles peuvent être aussi incluses dans la balise (exemple, <Biblio subject 1 = Brochures CFONB>) (*encadré 2*).

Comme on le verra plus loin de manière détaillée, un ensemble d'informations peut être relié aux balises utilisées pour un texte donné (intitulé, imbrication, ordre de succession, statut obligatoire ou facultatif) et constitue ce que l'on appelle la DTD (Définition type de document, *encadré 2*) de ce texte. Elle peut être soit directement déclarée dans le corps du document, soit fournie séparément. La DTD est toujours facultative en XML car elle n'est pas nécessaire si un document respecte les spécifications officielles XML ; elle est par contre indispensable si des règles particulières sont appliquées à un docu-

ment. Cet aspect facultatif donne une grande souplesse d'utilisation à XML.

Si les balises encadrent directement l'information qu'elles désignent (exemple, <prénom>karl</prénom>), il s'agit de données élémentaires, dans le cas contraire, <auteur> par exemple, qui encadre les nom et prénom, il s'agit de données composites. On parlera le plus souvent dans ce cas de *core components* en anglais. Elles peuvent avoir une hiérarchie, l'élément «Auteur» par exemple comprend d'abord le prénom (balise «prénom»), ensuite le nom (balise «nom»). Ainsi, à l'instar des messages Edifact, les documents XML ont une structure arborescente qui prend en compte la hiérarchie des données. Les règles auxquelles doit obéir cette structure définissent l'intitulé des balises, leur imbrication (hiérarchie), le caractère obligatoire ou facultatif des informations balisées ainsi que leur ordre de succession lorsqu'il y en a un. Ainsi «?» signifie que l'information est facultative dans le message ; «*» signifie qu'il s'agit d'un champ répétitif (plusieurs occurrences possibles) ; lorsqu'une information n'est pas accompagnée de l'un de ces deux symboles, cela signifie qu'elle est obligatoire.

La conjugaison de l'ensemble de ces caractéristiques a permis de définir un certain nombre d'éléments composites qui peuvent être réutilisés pour bâtir un message dont ils constituent une partie, voire constituer à eux seuls un message réutilisable dans plusieurs contextes.

LES BANQUES FACE À UNE ÉVOLUTION INÉLUCTABLE

XML offre donc une grande souplesse d'utilisation. La montée en charge de son utilisation dans les relations de type EDI entre les banques et leurs clients se justifie aussi par certains facteurs non né-

gligeables. Au plan technique, XML autorise la séparation des données et leur représentation, ce qui permet d'envoyer les mêmes données, voire des blocs de données, vers des terminaux d'affichage différents. Dans le même esprit, XML est un format qui permet de travailler avec des systèmes très hétérogènes, ce qui est très largement le cas dans les relations clients-banques. Au plan commercial et marketing, cette technique est fortement poussée par les fournisseurs de logiciels qui font petit à petit migrer vers XML le parc de leur offre. Il est donc patent que les banques auront à s'adapter à cette évolution. Ceci constitue une condition *sine qua non* pour que les banques, qui jusqu'à présent définissaient les standards dans ce domaine, ne soient pas mises en situation de subir des standards élaborés ailleurs. Au plan bancaire enfin, des acteurs d'importance, comme Swift et Bolero, ont pris la décision de privilégier XML pour leurs applications nouvelles et d'y faire migrer le parc existant. Il n'est donc guère envisageable que les banques ne suivent pas ce mouvement.

GÉRER LA COEXISTENCE DE PLUSIEURS LANGAGES

Ceci ne signifie bien évidemment pas qu'XML va remplacer l'Edifact, beaucoup d'applications existantes dans les entreprises vont continuer à fonctionner sur cette base pendant encore un certain temps, et les banques devront continuer à fournir les services en Edifact à leurs clients. Le même raisonnement vaut bien évidemment pour

les normes CFONB qui continuent à satisfaire un besoin important pour les entreprises. Les banques, pour répondre aux diverses attentes de leurs clients, vont donc être amenées à gérer aussi bien des formats fixes que variables, à structure arborescente ou à plat. Ceci implique, au

“ Un processus de simplification et d'harmonisation des langages utilisés a été engagé. ”

moins pour l'avenir proche, qu'XML ne constitue pour les banques qu'un format de plus à gérer, les autres continuant à être utilisés pour des volumes de transactions très significatifs.

La coexistence de plusieurs syntaxes peut donc d'ores et déjà être considérée comme un fait acquis. Mais il est aussi essentiel de préserver l'assiette commune, à savoir la définition des *business process* issus du processus de modélisation, les définitions et la sémantique des données, notamment l'acquit essentiel d'Edifact dans ce domaine. On ne saurait trop insister sur l'importance cruciale qu'il y a pour les banques à faire en sorte que les travaux sur les données utilisées par la sphère financière, qui sont menés dans différentes enceintes (Edifact, Edifact Working Group, Swift, Comité européen de normalisation bancaire), le soient avec un souci de cohérence maximale.

Ceci ne sera bien sûr assuré que si la communauté bancaire s'investit fortement dans ces travaux, soit directement soit par l'intermédiaire de Swift ou de l'EWG qui paraissent être les organismes qui réunissent les meilleures compétences sur ce sujet extrêmement important. ■

3. Où trouver les DTD ?



■ Les Définitions type de document (DTD) dites officielles et actuellement utilisées sont disponibles sur le site web <http://www.w3.org/TR/REC-xml>. Leur traduction en français peut être consultée sur le site http://babel.alis.com/web_ml/xml/REC-xml.fr.html.