

Le document dans un réseau de communautés d'intérêts : un avatar ?

Joël Gardes¹, Dominique Dionisi², Eric Trupin², Jacques Labiche²

¹ France Telecom R&D

² Laboratoire PSI FRE CNRS INSA et Université de Rouen

**joel.gardes@rd.francetelecom.com
Dominique.Dionisi@insa-rouen.fr
Eric.Trupin@univ-rouen.fr
Jacques.Labiche@univ-rouen.fr**

1. Introduction : évolution attendue de la communication interpersonnelle

Le concept de document a sensiblement évolué depuis la mise au point des techniques permettant sa dématérialisation. D'objet entreposé dont l'accessibilité ne dépend que des capacités de publication et de diffusion d'un éditeur, le document devient un objet de communication et d'échange.

Cela provient de la banalisation des outils de composition de documents et de transmission multimédia et multimodale d'information. Il ne s'agit plus de savoir comment diffuser le plus largement possible un volume important d'informations, mais plutôt d'un retour à une définition que l'on peut qualifier d'épistolaire du document, à savoir, pour un auteur communiquer à un tiers lecteur, une information. Cet acte de communication implique qu'un document doit pouvoir être trouvé ou retrouvé dans un réseau, lisible par un destinataire ou un groupe de destinataires et, surtout, compréhensible par ceux-ci.

La nécessité d'associer des critères physiques, logiques et conceptuels à la notion d'efficacité de l'acte de communication, implique que l'instrumentalisation de l'échange d'information au travers de documents doit être normalisée. Cette normalisation nécessite de définir précisément le système de communication dans lequel évolue l'information avec son cycle de vie ainsi que les différents objets circulant dans ce système, leurs propriétés et leurs relations. Un tel système de communication ne peut être vu comme un bloc monolithique si l'on souhaite prendre en compte la dimension "partage de connaissances".

On préférera alors le régionaliser en communautés d'intérêts, en nommant et typant les canaux et constituants de ces canaux, et en définissant pour chacun des canaux leurs propres ontologies. Ces ontologies ont pour discours non seulement les usages, les connaissances liées à ces usages, les mises en situation ou en contexte, mais également les modalités de représentation de l'information ainsi que les intentions et buts intervenant dans cette communication c'est à dire l'ensemble des connaissances donnant un sens aux objets invoqués dans un acte de communication.

Ces ontologies définissent ce que nous appellerons une infosphère et il s'agira alors de pouvoir mettre en œuvre un protocole d'interconnexion entre infosphères. En effet, que ce soit dans l'utilisation d'un formulaire, la co-rédaction d'un brevet, ainsi que dans tout acte faisant intervenir comme médium le document, nous serons toujours amenés à formaliser et décrire des traitements, des données et des contextes de mise en œuvre, dont la finalité sera de définir et d'exécuter un contrat de lecture entre une source d'information et un destinataire. La finalité d'un tel contrat sera de permettre des interactions, allant de la simple lecture jusqu'à la co-rédaction d'un document, qui guideront la reconstruction du sens par des utilisateurs à partir d'une certaine saillance des traces.

Une des conséquences de cette régionalisation d'un système de communication de documents circulants est la nécessité de chercher quelles sont les interfaces naturelles des infosphères. Du fait de la complexité et la variabilité des objets circulants (classes et méthodes), il semble intéressant de considérer chaque infosphère comme un réseau ouvert (une application « cognitive ») dont l'interconnexion sera assurée par des interactions spécialisées par type d'objet. D'autre part, ce processus d'interconnexion nécessitera des protocoles d'établissement de dialogues ou de transactions entre infosphères afin de rendre compatible les représentations de contenus aux usages intentionnels et attendus des utilisateurs. L'ensemble des protocoles mis en œuvre dans cette interconnexion portera sur des discours allant de la forme physique du document jusqu'aux jeux de métadonnées reflétant un usage sémantique sous la forme d'applications des objets du document.

Avec une telle organisation il devient possible de mettre en œuvre une approche système dérivée de l'interconnexion de réseaux de télécommunication, à savoir le

modèle ISO de l'OSI conçu pour permettre le dialogue d'applications informatiques qui sont des applications « cognitives » si on inclut dans le système les interactions avec les utilisateurs.

2. Evolution du statut du document dans une organisation

Le document a subi sa révolution avec l'apparition des systèmes informatisés. Même si le paradigme initial de la construction d'une documentation "zéro papier" demeure utopique, l'informatisation de la documentation a profondément fait évoluer le rôle du document dans une organisation quelle qu'elle soit. Cette évolution de l'utilisation de l'objet document a été induit non pas par la nature des informations transportées, mais par la diversification des supports. On peut introduire le néologisme de multimédiatisation du document succédant à sa dématérialisation.

On peut dire, pour résumer, que cette révolution s'est faite en deux temps.

Initialement, la conception, l'édition et la publication de documents était centrée sur le papier et nécessitait des postes exploitant des outils et méthodes d'impression (le métier de l'imprimeur), ainsi que des systèmes de classement ou d'archivage spécialisés (le métier de documentaliste/archiviste). La production d'information passait inéluctablement par ces postes qui, du fait de leur spécialisation, étaient centralisés au sein d'une organisation. Corollairement, la circulation d'informations par des documents se faisait dans une logique de diffusion d'un producteur vers un consommateur de documents. Le producteur avait toutes les facilités et surtout, l'expertise accessible, pour agir sur le contenu même du document. Le consommateur, quant à lui, ne modifiait pas le contenu mais pouvait l'enrichir par un jeu d'annotations et de documents associés dont la réelle prise en compte passait par la recomposition du document initial.

La première révolution a été marquée par l'arrivée des traitements de texte et autres éditeurs graphiques qui ont permis, dans un premier temps, de désanctuariser les fonctions de composition de documents, tout en conservant l'édition papier comme support unique. Cela a conduit à organiser la circulation de documents dans une logique de multidiffusion. Si le producteur de document, au sens de la première étape de notre révolution, conservait ses prérogatives de gestion globale de la documentation de l'organisation ou de l'entreprise, les membres de cette organisation ont eu la faculté de devenir eux-mêmes, producteurs de documents. Par contre, du fait de la persistance d'une logique de diffusion dans l'organisation de la circulation d'information, associée à la persistance du médium papier, le document conservait des propriétés que l'on peut résumer ainsi :

- Le contenu conservait un caractère pérenne car il était difficilement modifiable du fait de la persistance de l'étape de recombinaison.
- Une charte de représentation de l'information, propre à l'organisation pouvait être facilement appliquée par l'utilisation de feuilles de styles
- Le document circulait dans une organisation tel quel sur différents supports tel que le papier ou des mémoires de masse (disquettes, bandes, etc...) qui ne permettaient pas d'interaction instantanée entre le producteur et le consommateur. Par conséquent, l'intégrité du document initial était préservée.
- La dématérialisation ne se comprenait que par la numérisation sous la forme d'image de documents papiers à des fins d'archivage reprenant un plan de classement identique à celui du document papier.

La deuxième révolution est celle que nous vivons à présent. Elle a été lancée par la banalisation des réseaux de téléinformatique. En plus des systèmes de traitement de texte, de PAO (Publication Assistée par Ordinateur) devenant facilement appropriables par des utilisateurs non avertis, les stations de travail ont été interconnectées en réseau ce qui a bouleversé la temporalité du traitement des documents. D'une logique de diffusion d'un producteur vers plusieurs consommateurs, nous sommes arrivés à une logique où chaque élément du réseau de circulation d'informations est à la fois producteur et consommateur de contenus numériques et non plus seulement de documents. La conception et la production de documents est devenu interactive, collaborative et les fonctions d'édition/publication sont passées au second plan. Dans ce contexte, le document n'est plus un objet pérenne pour une organisation (à moins qu'on l'archive sous forme papier ou d'images numérisées), mais un conteneur de données dans lequel chaque nœud du réseau de l'organisation peut intervenir pour modifier à la fois la structure et la nature du contenu. Une information, dès lors, peut avoir plusieurs représentations (images, textes, dessins) et plusieurs formulations pouvant même aller jusqu'à induire des formes contradictoires susceptibles de pénaliser la valeur même de l'information. Par conséquent, dans cette dernière révolution on s'aperçoit qu'un élément fondamental est la remise en cause de la valeur même de l'information. Celle-ci ne peut plus être absolue mais doit être relativisée par rapport à deux critères : le support utilisé et la sphère d'information qui l'a produite. Il devient dès lors, indispensable de faire évoluer les modèles d'organisation intervenant dans la conception des systèmes d'informations documentaires en prenant en compte cet aspect d'interconnexion forte entre sphères d'informations personnalisées.

3. Traitement numérique du document

Nous référant au travail collectif de réflexion sur la notion de document au sein du Réseau Thématique Pluridisciplinaire 33 du département STIC du CNRS [PED03], nous pouvons considérer le document comme un objet de communication régi par des règles de mise en formes plus ou moins explicites, qui matérialisent un contrat de lecture entre un auteur et un lecteur. De ce point de vue, le document peut être observé comme un canal de communication (figure 1) qui lie ces deux acteurs au travers de la mise en place d'un protocole de communication qui aide d'abord à transcrire sur un support de l'information qui sera ensuite extraite par le lecteur.

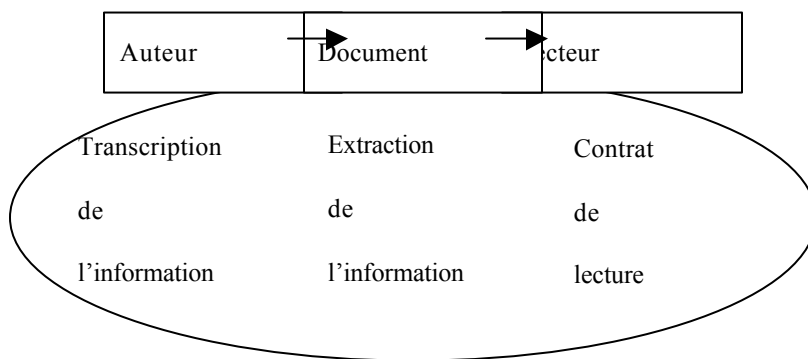


Figure 1 : le document vu comme un canal de communication

Ce protocole définit un ensemble de règles (topologiques, typologiques, syntaxiques) et de contraintes sémantiques qu'il convient de modéliser si l'on veut mettre en place un système d'interprétation automatique de document, afin d'en extraire des informations liées à un domaine d'usage du document et selon des attentes (requêtes) d'un utilisateur appartenant à une communauté d'intérêt. Ces informations respectent une organisation logique propre à l'intégrité du document initial mais forment également des unités d'informations indépendantes, repérables par des métadonnées et personnalisables pour un lecteur qui dispose de son propre référentiel modélisable par une ontologie.

Du point de vue des spécialistes du traitement d'image de document qui s'attachent à construire des systèmes d'interprétation automatique de ces images, il est indispensable de doter le système de connaissances lui permettant d'extraire l'information qu'elles supportent. Ceci s'inscrit dans un chaînage Données-Information – Connaissance [DEN03] dans lequel la connaissance détenue par le système permet d'extraire l'information supportée par les documents à partir des données image. Ces connaissances traduisent les éléments du protocole de communication défini autour du document mais doivent également exprimer les

intentions des auteurs du document et des utilisateurs (lecteurs) du système d'interprétation dans un cadre défini. On peut distinguer trois types de connaissance selon [NEW82]. Les connaissances opératoires sont liées à la paramétrisation des outils de traitement d'image et à l'évaluation de leurs résultats. Elles traduisent le savoir-faire du spécialiste de traitement d'image. Les connaissances descriptives concernent la modélisation des contenus de document en construisant des modèles. D'autre part, elles décrivent aussi le domaine d'usage en définissant des contraintes sémantiques. Enfin les connaissances stratégiques visent à planifier les scénarios de traitement pour répondre aux attentes de résultat des utilisateurs. L'intégration de ces connaissances fait apparaître clairement la nécessité de doter le système de capacité d'interaction de différents acteurs avec celui-ci : Ces acteurs, chacun relié à son propre réseau, service ou organisation, sont :

- Le concepteur du système et/ou le spécialiste en traitement d'image. Il intègre les connaissances descriptives, opératoires et stratégiques liées à l'interprétation des images de document.
- Le spécialiste du domaine d'usage, ou le concepteur du document. Celui-ci détient la connaissance liée au domaine d'usage, ou au métier, et produit des documents pour l'utilisateur. Il dialogue avec le concepteur dans une phase de spécification du système ;
- L'utilisateur est en interaction avec le système et peut rendre compte au concepteur des défauts du système comme un retour d'expérience. Il est muni d'une intention sur le document pour déclencher une action.

Mais plus généralement, le concept de connaissance couvre différents aspects, et selon les points de vue, il peut correspondre à des données observées, des informations structurées, un mécanisme de raisonnement où à des processus d'analyse. De plus la connaissance peut provenir d'un opérateur humain, de systèmes d'information, de bases de données, d'expériences ou de documents par exemple. L'ensemble des connaissances relie les informations et les savoirs [DUP99], mais aussi et surtout les acteurs pour lesquels la connaissance est indispensable pour préparer une décision et pour agir. De ce fait, ils sont tout à la fois porteurs, destinataires et utilisateurs de cette connaissance en vue de l'action, et pour l'action. Ainsi la question est de savoir comment exprimer la connaissance, l'identifier, la représenter et la partager. D'autre part, comment les acteurs peuvent-ils acquérir la connaissance et se l'approprier ? Il faut tout au-moins rendre accessible l'information à la bonne personne et au bon moment. Dans le contexte des documents, cette information est ce que les systèmes et/ou acteurs, producteurs et consommateurs de documents, s'échangent. L'information écrite ou graphique, médiatise une pensée, se matérialise et laisse une trace sur un support physique ou numérique.

Une autre analyse montre que deux types de connaissances co-existent selon [NON94] : les connaissances explicites et les connaissances tacites. Les premières sont saisies et stockées dans des archives, des systèmes de gestion de document, des bases de données ou des bibliothèques. Elles sont accessibles, transférables et formalisées. Les secondes sont très suggestives. Elles comportent des schémas mentaux, des expertises et des compétences dans des contextes spécifiques, et on peut très difficilement les formaliser. Elles ne sont pas explicitées. Par contre, l'utilisateur peut construire des schémas mentaux, c'est-à-dire des connaissances tacites qui lui sont propres, à force d'utilisation du système et donc d'apprentissage également. De cette manière, il pourra (ré)agir selon son intention en fonction de l'information que lui transmet le système. Mais cette information n'existe que dans le cas où l'utilisateur possède ces connaissances.

Ces éléments montrent que lorsque les documents sont liés à un domaine d'usage et/ou une communauté d'intérêt si l'on s'inscrit dans une organisation (l'usage peut s'inscrire dans une pratique professionnelle et donc correspondre à un métier), la conception d'un système de traitement numérique de document « métier » constitue alors de réels enjeux scientifiques car le document n'est plus considéré en tant que tel, mais du point de vue des stratégies d'acteur mises en place autour de lui, et au sein d'une organisation. Pour répondre au besoin de modélisation et d'intégration des connaissances mises en jeu, il s'agit alors d'analyser les contraintes qui font du document un élément déclenchant dans un contexte donné au sein d'une sphère d'activité (infosphère), les normes et les pratiques sociales en vigueur, les cycles de vie du document et la typologie des corpus en fonction de leur degré d'interaction avec l'organisation considérée. Les auteurs et les lecteurs, socialement situés dans l'organisation sont au centre de ce dispositif et il s'agit principalement d'analyser les interactions entre un corpus de documents et la communauté d'intérêt à laquelle ils appartiennent.

4. Une vision de l'organisation par les infosphères

Nous avons vu qu'au sein d'une organisation, la désanctuarisation des entrepôts d'informations se traduit par le fait que chaque entité (i.e. chaque utilisateur) est potentiellement à la fois source et destinataire d'informations. Aussi, dans la conception et l'architecture des systèmes d'informations il devient indispensable de prendre en compte la personnalisation de l'information et de donner au système une vision de l'utilisateur au travers de modèles.

D'autre part, nous verrons que le réseau assurant l'échange d'information devient un élément actif du système car il ne devra pas seulement se contenter de mettre en relation les utilisateurs ou groupes d'utilisateurs mais également assurer les fonctions

d'intermédiation permettant de personnaliser l'accès et la représentation de l'information.

Enfin, la proposition de prendre en considération la personnalisation implique que le système d'information devient un système ouvert, avec la possibilité d'inclure de nouveaux membres dans une organisation et cette propriété offre un terrain très favorable à l'interconnexion de systèmes.

Cependant, si la représentation de l'information tend à privilégier la personnalisation, il faut se poser la question de la structuration même du système en termes d'urbanisme et d'architecture de manière à ce que celui-ci soit capable de fédérer des "niches d'informations" ou "infosphères". Il devient donc important de définir l'objet "infosphère" lui-même en tant qu'entité homogène et, pour que le système lui-même reste homéostatique et intègre, voir comment les interfaces et les protocoles d'échanges entre infosphères peuvent être organisés.

4.1. *La notion d'infosphère*

L'utilisateur, au sein d'une organisation ou amené à échanger de l'information avec une organisation, possède son propre environnement matériel, logiciel, ainsi que son propre savoir, ses propres croyances et habitus. C'est avec cet environnement que l'utilisateur s'est construit sa vision du monde. Dans les systèmes d'informations "classique" la recherche d'information nécessite de la part de l'utilisateur un investissement cognitif certain passant par l'apprentissage des modes de transaction et l'appropriation de la perception de l'information imposée par le système. Ainsi l'homogénéité du système est donnée par l'homogénéité de la représentation de l'information avec pour conséquence une remise en cause du capital cognitif de l'utilisateur si le système évolue.

Or, nous avons vu que l'interconnexion massive de sources d'informations limite, voire interdit, le principe de représentation unique de l'information. D'autre part, l'environnement réseau actuel favorise cette interconnexion. Il s'agit là d'une rupture technologique majeure qui induit des ruptures de services et d'usages importants.

En terme de rupture de service, on se place dans le contexte des très grandes bases de données avec l'apparition de représentations de l'information conférant à celle-ci un caractère imprécis, contradictoire, voire des données absentes. Chaque nouveau service se devra d'inclure des fonctionnalités de gestion de l'ambiguïté.

En terme de rupture d'usage, on assiste à une forte émergence de l'interaction multimodale qui pose des questions sur les "valeurs perceptives et cognitives" de l'information perçue par l'utilisateur. On doit se poser la question suivante :

l'environnement de l'utilisateur lui permet-il de lire et de comprendre l'information reçue ? D'autre part, si l'on désire maintenir de bonnes conditions d'acceptabilité des services, le système (ou "réseau d'interconnexion") et les environnements immédiats de chaque utilisateur doivent pouvoir évoluer séparément. Le système ne gère plus les utilisateurs mais est amené à fédérer des "bulles d'informations personnelles" en assurant toutes les fonctions de mise en relation, de synchronisation des processus entre applications et de représentation de l'information. Chaque "bulle" devient donc ce que l'on appellera une infosphère. Dans cette ébauche de modèle, l'infosphère devient l'unité architecturale de base des systèmes d'information.

4.2. *L'organisation : un regroupement d'infosphères*

Le modèle des infosphères est intéressant pour définir des modes de communication point à point et multimodaux. En effet, il différencie bien la partie locale de traitement de l'information propre aux utilisateurs de la partie communication de l'information. Nous allons essayer d'étendre ce schéma à une organisation en prenant comme exemple, une entreprise.

Une entreprise est une organisation qui a un but, produire. Cette production peut être matérielle (construction d'objets à vendre) ou immatérielle (vente de services). Afin d'accomplir ce but commun, l'entreprise s'organise en processus qui peuvent se différencier en deux grandes catégories :

- Les processus externes comme la relation clientèle, les achats/commandes
- Les processus internes comme la gestion du personnel, des stocks, de la qualité

D'un point de vue "ontologique", on s'aperçoit en fait qu'une entreprise n'exploite jamais un seul métier. Nous aurons des techniciens, des juristes, des comptables et d'autres corps de métier ayant tous un même but : produire.

L'information, quant à elle, concerne à la fois le but de l'entreprise et la gestion des moyens pour atteindre ce but, ainsi que les dérives sur les couples objectifs/moyens.

Il est parfaitement possible de modéliser les données et processus de l'entreprise comme cela est proposé dans les systèmes de gestion intégrés de la famille des ERP (Entreprise Ressource Planning). Cependant, la mise en place de ces solutions intégrées passe toujours par des étapes préliminaires d'audit et de réorganisation de l'ensemble des processus et données de manière à se mettre en conformité avec les modèles d'organisation prévus par ces solutions. Dans le contexte de grandes entreprises ou d'organisations ayant plusieurs buts (par exemple, vente de services et

de produits), souhaitant inclure ces solutions intégrées dans leurs systèmes d'informations, ce travail d'initialisation correspond à des postes lourds sur les plans budgétaires et de gestion des ressources préexistantes.

Le fait de choisir un logiciel intégré de gestion implique donc une évolution parallèle du système d'information et des environnements utilisateurs, et la difficulté de mettre en place ce genre de système réside dans le fait que les infosphères ne sont pas bien identifiées actuellement.

Une solution alternative à moyen terme semble possible si l'on applique le modèle des infosphères. L'hypothèse de départ est simple. Chaque processus de l'entreprise se déroule dans une infosphère dans laquelle travaille une communauté d'intérêt ayant localement sa propre ontologie (son métier, ses intentions et ses buts propres) et amenée à dialoguer avec les autres processus de. Dès lors, l'organisation du système d'informations de l'entreprise deviendra une interconnexion d'infosphères.

Sur le plan technologique chaque utilisateur devient un dépositaire de fichiers numériques et le système, à partir d'une requête faite à l'aide d'un moteur de recherche, reconstruit un résultat de requête. Nous sommes dans une logique de "gopher" tels que nous les avons connus dans les années 1980. Cela veut dire que l'on peut concevoir d'ores et déjà un système de fichiers (le niveau le plus bas d'un système d'information) totalement décentralisé dans lequel chaque utilisateur est à la fois dépositaire et demandeur d'informations.

En extrapolant la même organisation pour les interfaces utilisateurs et outils de traitements, nous arrivons à la définition de véritables systèmes d'informations virtuels dans lesquels chaque infosphère aura une interface normalisée permettant d'assurer toutes les fonctions de mise en relation, de synchronisation et de représentation personnalisée des données.

Il convient, de se pencher sur la manière dont les différentes infosphères vont dialoguer et définir, en particulier, cette fonction d'intermédiation, apparue du fait que l'on applique un postulat de maintien de la personnalisation (singularisation) dans les infosphères.

La question de l'interconnexion aboutit à la conception de l'interface des infosphères. Quel sera son rôle, quel genre de messages va t'elle gérer et avec quel langage ?

Si l'on se focalise sur les buts de ces interconnexions, en affirmant que l'interconnexion entre infosphères est inféodée à des buts pérennes (durablement appris) au sein d'une organisation et des buts "volatils" (fortement contextualisés par les intentions) dans le cas d'échange d'informations entre organisations, on ne remet

pas en cause la nature des infosphères ni de leur interface. Seuls des critères temporels servent à typer les infosphères.

Le regroupement au sein d'une organisation induit une mémoire stable du système qui permet de minimiser les étapes d'établissement du dialogue qui concerneront la mise en relation, la synchronisation et la médiation de représentation de l'information. Alors qu'une interconnexion temporaire, comme par exemple, une démarche administrative (comme celle abordée avec l'étude sur le RMI) nécessitera un établissement de dialogue fortement interactif, correspondant à la négociation entre le demandeur chargé de remplir un formulaire qu'il devra avoir lu et compris et l'agent administratif qui devra expliciter et assurer la cohérence et la complétude de la demande par rapport à une réglementation.

Les indices qui permettront dans chacune de ces deux infosphères de savoir si l'échange d'information respecte ces critères de conformité ne portent pas seulement sur les traces portées sur le document, mais également par un partage de connaissances reformulées entre la source et le destinataire. Or, à l'issue de cette transaction de dialogue, il ne restera plus que des traces sur le document. Par conséquent le résultat de ce dialogue n'est pas une information mais un signe qui fera l'objet d'une lecture et d'une interprétation au niveau de l'interface. Ceci nécessite de revenir sur le modèle même de l'information, classiquement représenté par une dyade objet-sens. En effet, ce modèle ne rend absolument pas compte de ce qui reste le seul artéfact transporté par le document, à savoir des traces codées sous la forme de signes.

5. Un modèle de représentation de l'information

Il apparaît donc nécessaire de revenir sur le modèle de représentation de l'information afin de pouvoir prendre en compte les questions soulevées par la personnalisation de la représentation de l'information et de la définition des infosphères et de leurs interfaces. Ces questions portent essentiellement sur la multireprésentation d'une information : celle-ci peut être représentée soit sous la forme de documents électroniques, soit d'images (fichiers électroniques). Et, à l'intérieur de ces "conteneurs de données", nous trouvons des signes qui peuvent être soit du texte typographié, soit des traces manuscrites, soit toute forme d'artéfact. Se pose alors le problème de la multiinterprétation de ces artéfacts par différentes communautés d'intérêts amenées à dialoguer sur un même but.

Communément, on admet qu'une information est représentée par la dyade objet-sens. Dans un système d'information où l'on a une convention unique de représentation des signes, en l'occurrence, l'ASCII ou l'Unicode, ce modèle est

parfaitement applicable. Par contre, le fait d'introduire le document comme médium multimodal de communication d'informations, nécessite de faire évoluer cette représentation de l'information en incluant la notion même de signe.

Par conséquent, nous avons là un modèle basé sur une triade sémiotique et nous nous proposons d'appliquer un modèle phénoménologique de représentation de l'information, celui donné par la sémiotique de Peirce. Schématiquement, cette triade se compose :

- D'un objet : c'est un container de données telles qu'elles sont implémentées dans le système d'information ;
- D'un signe : c'est la matérialisation perceptible de l'objet, sa représentation au travers duquel l'utilisateur peut interagir. En l'occurrence, cela sera une typographie, un tracé manuscrit, un dessin, etc. ... ;
- Un interprétant : c'est l'ensemble des spécifications, des descriptions donnant à un objet son sens a priori ou son sens contextuel. Ce n'est pas le sens même du phénomène, mais les moyens et instruments permettant d'accéder au sens.

Ces trois éléments peuvent être instanciés selon l'usage et l'application et permettent de rendre compte assez facilement de la multireprésentation de l'information. Nous verrons dans ce qui suit que l'on trouve une cohérence forte entre cette représentation, somme toute très classique sur le plan épistémologique de l'information et les propriétés de la fonction d'intermédiation. Ceci nous permettra de jeter les bases d'une normalisation des échanges entre les infosphères.

5.1. Normaliser des échanges entre infosphères

Normaliser un processus implique de définir un modèle qui permettra de donner à un système une perception ou un point de vue sur l'information et sur les constituants ou instruments servant au traitement de cette information.

En ce qui concerne le dialogue entre infosphères, nous sommes en présence d'un acte de communication faisant systématiquement intervenir des actes de capture ou de perception et des actes de restitution d'informations. Aux extrémités de ce canal, se trouvent une source d'information et un destinataire. Il s'agit à présent, de repérer les différents processus et constituants associés intervenant dans tout acte de communication. Il nous a paru intéressant de revenir sur le modèle du canal de transmission de Shannon qui est, encore actuellement, à la base de tout système de télécommunication. La raison de ce choix est l'existence de propriétés fortement structurantes pour un système de communication, associée à un grand degré de

généricité. D'autre part, nous verrons que ce modèle classique introduit des définitions des artefacts transporté par le canal qui se révèlent applicables aux échanges entre infosphères.

5.2. *Le canal de communication de Shannon*

La finalité du modèle de canal de communication tel que l'a introduit Shannon est de proposer un processus générique de conversion de l'information en vue de son transport. La règle de base est que la transmission doit être adaptée au support. Très schématiquement, nous trouvons une source d'information et un destinataire aux extrémités de ce canal.

Nous allons rappeler les composants de ce modèle en prenant la transmission dans le sens de codage de l'information. Le canal étant symétrique, nous aurons exactement les fonctions inverses à l'autre extrémité.

La source émet donc de l'information. Cette étape est assurée par des transducteurs.. Le résultat de cette conversion est un flux de messages ayant comme propriétés :

- d'être conformes à la source,
- de n'avoir aucune sémantique propre, sauf dans le cas où le système prévoit un codage de la source pour inclure le paramétrage du transducteur dans le processus de communication.

Il s'agit d'une adaptation physique de l'information.

L'étape suivante correspond à l'émission des messages après codage dans le canal. Cette opération est dévolue à l'émetteur (ou au codeur en transmission numérique). A l'issue de cette opération, nous obtenons des signaux codés transportables dans le canal.

Le plus important est la nature même du codage, qui correspond à une conversion logique de l'information. Il s'agit de transporter une description conforme des messages dans le canal. Cette description est en fait un langage correspondant à une convention établie entre les équipements d'extrémités du canal. Les spécifications de ce langage font référence à une morphologie, une syntaxe et une sémantique qui garantissent, en fonction des spécifications physique du canal, la conformité de transmission de l'information.

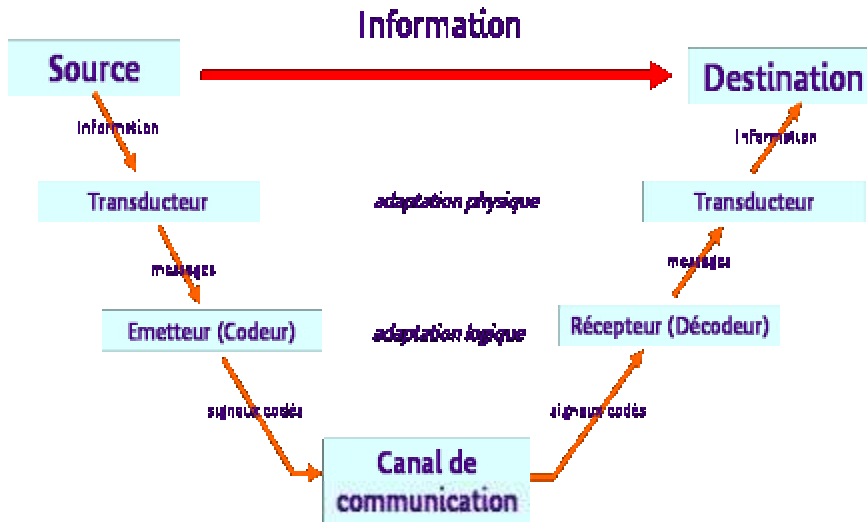


Figure 2 : canal de communication

L'échange d'information entre infosphères peut être assimilé à un acte de communication. Il est donc intéressant de voir s'il est possible d'appliquer, non pas, le modèle de Shannon en tant que tel, mais sa philosophie.

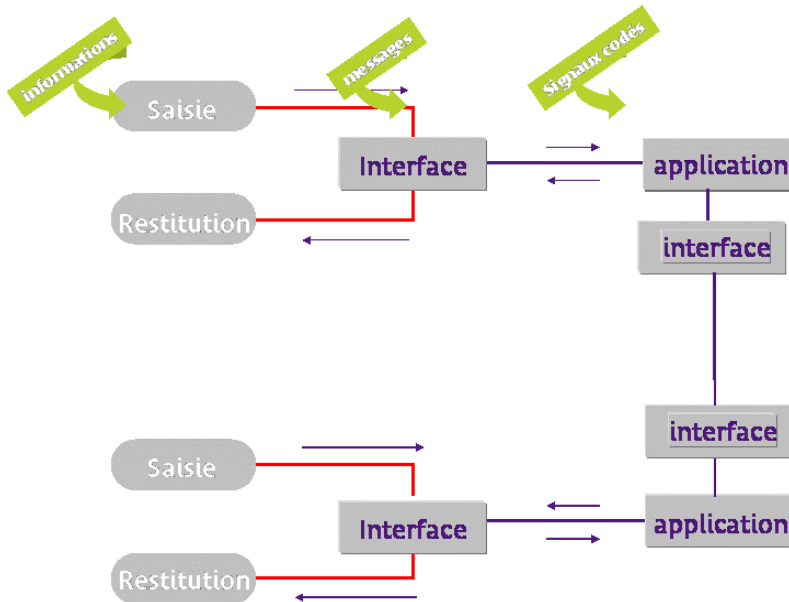
Le problème de départ, à savoir transmettre de l'information entre infosphères hétérogènes, que celles-ci appartiennent à une organisation ou non, met en place, à nouveau, une source d'information et un destinataire. Dans l'notre exemple, la source dialogue avec une application et c'est une application qui retransmet l'information à un destinataire. L'application joue donc le rôle du canal de transmission.

Quelle que soit la modalité de transmission de l'information envisagée, nous allons trouver un dispositif matériel pour la capture de cette information et un dispositif pour la restitution. Pour la capture, il peut s'agir par exemple, du clavier, mais également d'un couple stylet/dalle tactile, d'une caméra ou d'un scanner. Le rôle de ces dispositifs de capture est de générer des messages, qui ont comme propriété surprenante de ne faire aucune référence à la sémantique de l'information. La restitution quant à elle sera assurée par un écran. Son rôle sera de présenter une forme perceptible de l'information par l'utilisateur. Les dispositifs de restitution reçoivent des messages, également dépourvu de sémantique en relation avec l'information transmise. Les messages sont soit des tracés, soit des images bitmap, soit des couples d'impulsion électrique et peuvent, en fait, être assimilés à des stimuli. La construction de l'information et de son sens, à partir de messages se fait par un dialogue interactif entre l'utilisateur et le système, c'est pour cette raison que le modèle proposé associe systématiquement comme équipements d'extrémité un

dispositif de capture et un dispositif de restitution. Ce n'est qu'à partir des messages que l'utilisateur peut construire sa vision sémantique de l'information, charge au système de gérer des messages saillants.

Les dispositifs de capture et de restitution, parties matérielles de notre système d'interconnexion, sont reliés à une interface logicielle. C'est elle qui reçoit ou émet les messages d'une part, et assure, à partir des spécifications d'une application, le codage de ces messages en signaux. Si, par exemple, l'artéfact saisi est un tracé vectoriel, nous trouverons une interface qui aura en charge de coder ce tracé en encre électronique (par exemple, en InkML). Cette encre électronique codée sera ensuite transmise à une application.

On s'aperçoit ainsi que le seul «sens» perçu par le système est qu'il s'agit d'une encre électronique dont le codage a été fait avec des règles morphologiques, syntaxiques et sémantiques propres à la nature du tracé et non au sens de l'information. Pour arriver au sens de l'information, nous devons reconnaître l'encre saisie pour donner à l'information son interprétant (le sens du texte que l'on aura écrit), mais également être capable de convertir la représentation de l'information (les applications de traitement de l'information ont besoin de l'ASCII pour manipuler les objets).



Par rapport à la proposition de représenter l'information sous la forme d'une triade sémiotique, on s'aperçoit qu'en partant du canal de Shannon, nous retrouvons les éléments composant cette triade :

- Les objets : ils correspondent aux données structurées manipulées par les applications. Bien entendu, une information peut être composée de plusieurs objets plus ou moins dépendants.

- Les interprétants : ils correspondent aux spécifications de codage des messages. Nous voyons ici que le sens n'est pas direct mais dépend de spécifications données par les applications.

- Les signes : ce sont les messages de notre canal émis et perçus par les dispositifs de saisie/capture et de restitution

L'information peut avoir des formes très variées. Mais systématiquement, il conviendra de définir son instanciation sous la forme de triades indissociables, quelle que soit l'application envisagée. De l'instanciation en signes dépendront le choix des dispositifs matériels. Les interfaces, quant à elles, seront tributaires de l'instanciation en interprétants des messages et le lien avec les objets sera le siège des applications.

La canal ainsi formé permet de transporter une instance d'information de bout en bout. Maintenant, que se passe-t'il si le processus d'échange d'informations nécessite l'interconnexion d'infosphères hétérogènes?

Pour illustrer cette question prenons le cas d'une saisie d'information manuscrite à inclure dans un traitement de texte. Nous avons là deux applications, deux canaux en fait, car l'application de saisie d'une encre électronique et le traitement de texte sont des objets susceptibles de fonctionner indépendamment l'un de l'autre.

En terme de représentation de l'information, une des applications ne sait gérer que des encres électroniques, l'autre application, quant à elle, attend de l'ASCII, même si elle permet de gérer des représentations de caractères variées (les polices). Nous sommes donc dans un contexte de représentation hétérogène de l'information que nous allons résumer ici comme étant un mot saisi. Afin de représenter ce mot de manière à ce qu'il puisse être pris en compte dans un traitement de texte, il va falloir le "reconnaître". En fait, il va falloir pouvoir passer d'une instance de représentation telle que prévue dans l'application de saisie d'écriture manuscrite, à une instance de représentation compatible avec notre traitement de texte. Dans l'exemple du mot, nous aurons un objet (le mot lui-même) et deux familles de signes (des tracés ou des caractères) et deux familles d'interprétants (le codage InkML et le codage ASCII). Par conséquent, nous devons mettre en place un protocole de conversion qui

s'assimile à un transcodage de l'information, à l'instar de ce que nous observons quand nous interconnectons des réseaux de télécommunication.

Ce transcodage fait apparaître qu'il s'agira d'une application particulière, capable de gérer les deux points de vue d'un même objet. Il s'agira donc d'organiser les spécifications de protocoles. Nous proposerons plus loin de s'inspirer des modèles d'interconnexion de systèmes ouverts, tels que spécifiés par l'OSI.

Mais, d'ores et déjà, nous pouvons dire que cette opération de conversion de protocoles correspond à l'établissement d'un contrat de lecture entre infosphères. En fonction de la portée de ce contrat, à savoir la possibilité offerte ou non d'interagir avec l'information transmise, on induit un nouveau type de bande passante qui rend compte de la complexité et la richesse de l'information échangée. Nous la nommerons "bande passante informationnelle".

5.3. *Notion de bande passante informationnelle*

Elle est définie à partir de l'interaction offerte via le contrat de lecture entre l'information affichée et un service utilisateur et/ou consommateur. Cette bande passante est un critère important d'interopérabilité des données entre infosphères, avec la question de leur représentation. L'échange de données peut se faire sur des supports allant de l'image (document numérisé) à une structure XML complète (XSL, DTD, RDF communs ou règles de conversion statiques connus).

Selon le degré de compatibilité entre les infosphères, il y a une fonction d'adaptation des contenus numériques et de leurs descriptions (métadonnées) à faire lors de l'interconnexion de processus. Cette fonction est notre fonction d'intermédiation. Elle doit fournir des processus de conversion intelligente de données allant de la rétroconversion de documents numérisés à la réingénierie de bases de données, en passant par la production de données exploitables par les systèmes d'indexation.

Cette bande passante correspond donc à l'acceptabilité et l'utilisabilité de l'information. C'est elle qui donne la valeur réelle à l'information. Par conséquent, le calcul d'un éventuel coût d'une transaction entre infosphères devra prendre en compte la largeur de bande passante informationnelle utilisée. Cette approche économique de calcul de la valeur de l'information est probablement une évolution forte à attendre des modèles d'affaire du web. En effet, cette approche ne prend pas le volume de données, ni le temps d'interconnexion, mais le "contrat de lecture" qui spécifie les conditions d'utilisabilité offertes et attendues par deux infosphères.

5.4. *Normalisation par les données ou par les interfaces ?*

Les données personnelles sont la propriété de celui qui en a fait la composition et l'acquisition (avec la restriction d'un usage propre ou personnel!) L'utilisateur développe ainsi une infosphère en organisant ses données selon ses propres modèles et ses modes de représentation: images, textes, dessins, multimédias permettent le plus souvent de décrire le même objet.

Par contre, il est certain que l'on ne doit pas considérer cette bulle d'information comme un système fermé. L'utilisateur sera toujours amené à communiquer des données et à en recevoir. On parlera alors d'interconnexion de ces infosphères. A partir de là, deux questions importantes se posent :

- • Comment assurer l'interconnexion de ces infosphères ?
- • Comment prendre en compte les points de vue des utilisateurs en terme de représentation des données ?

Dès que l'on parle d'interconnecter des Systèmes d'Informations, la question de la normalisation des échanges se pose. L'interconnexion des infosphères pourrait être obtenue par une normalisation forte des données conduisant à la définition d'un format pivot universel d'échange reprenant, par exemple, la philosophie des systèmes MPEG. Mais cette approche gestionnaire de l'information présente deux inconvénients majeurs principalement induits par la forte hétérogénéité des données :

- Du point de vue du *réseau* : une interdépendance entre l'infosphère et le réseau ferait que l'un ne pourrait plus évoluer de manière indépendante de l'autre, et le point de vue réseau serait privilégié au détriment du point de vue de l'utilisateur. ?
- • Du point de vue de *l'utilisateur* : une vision trop complexe des données (trop de formats, trop de représentations) le conduirait à *rejeter* à court terme tout service d'interconnexion personnelle.

Par conséquent, si la normalisation des données ne peut être faite, c'est peut être du côté des interfaces que se situe la réponse. Prenons l'exemple d'un service de "carnet de voyage électronique" dans lequel l'utilisateur souhaiterait récolter des données fortement hétérogènes à partir de différents canaux allant de son appareil-photo numérique (bientôt communicant) à des prospectus électroniques présents sur différents sites web, en passant par des annotations saisies au clavier ou ... manuscrites.

Dans un tel contexte, la question est : comment organiser les fonctions d'interfaçage susceptibles de rendre compatibles ces différents flux de données avec « l'infosphère »? Une forme d'intermédiation doit donc être assurée du côté du réseau d'interconnexion qui doit passer d'une logique de gestion des données à une logique de fédération des infosphères, définissant pour cela des services d'interface de conversion ou de reformulation des contenus auxquels l'utilisateur pourra souscrire. Cette approche a le mérite de privilégier le point de vue de l'utilisateur par rapport à celui du réseau, le conduisant normalement à accepter plus volontiers l'interconnexion devenue pour lui quasiment transparente.

5.5. Conclusion

L'ensemble des concepts que nous venons d'exposer ont pour base deux modèles classiques de communication. A savoir, l'approche phénoménologique de l'information en prenant en compte un modèle sémiotique incluant la représentation en signes et la décomposition du processus de communication d'information dans un modèle s'inspirant du canal de transmission de Shannon.

Il est intéressant de constater une certaine complétude entre ces deux modèles que l'on peut exprimer par des relations entre les composants du canal et les artéfacts transportés qui font directement référence au modèle triadique.

Ensuite, ces concepts ont un rôle structurant important dans tout système d'échange d'information hétérogènes : l'intermédiation, qui inclut des fonctions de reconnaissance d'artéfacts est située de manière stable dans notre système, quel que soit la modalité concernée.

Cet ensemble structurant permet de concevoir un environnement dans lequel chaque utilisateur ou groupe d'utilisateurs a la possibilité de gérer ses informations et traitements locaux indépendamment des spécifications du réseau d'échange. Ces infosphères sont donc des systèmes autonomes mais ouverts pour peu que l'on normalise leur interface de dialogue avec leur "monde extérieur". Sur le plan épistémologique, ces concepts peuvent faire référence aux théories sur la sémiologie (Peirce, Bateson) pour la reconstruction de l'information, tout comme aux théories sur l'inféron (Jacob) comme constituants d'un système complexe (Piaget) dont les infosphères en seraient un avatar.

Enfin, sur le plan économique, que ce soit pour la détermination de la valeur de l'information qui n'existe que si elle est associée à un contrat de lecture ou la personnalisation de la représentation de l'information qui nécessite de prendre en compte sa dimension multimodal, nous rencontrons là des enjeux importants de

l'économie du web. La vision en infosphères que nous proposons, commence à apporter des éléments de réponses.

Parmi ces éléments, intervient l'aspect normalisation des échanges. Nous avons vu qu'il n'est pas suffisant de spécifier un format pivot de format des données, étant donné que ce formalisme aurait une complexité combinatoire dans une hypothèse d'interconnexion point à point et maillée d'infosphères. Nous proposons plutôt de fédérer ces échanges en les normalisant par des spécifications de protocoles au niveau des interfaces, à l'instar des réseaux de communication ouvert. Ces spécifications, reprennent la philosophie OSI en s'organisant en couches de représentation des données allant de leurs instances physiques à leurs instances applicatives ou sémantiques. Le tableau ci-dessus correspond à une première hypothèse pour le document.

Couche	Exemples de données (non exhaustif)	Informations
applicative Traduit la finalité de l'objet	Ensemble cohérent de données s'inscrivant dans le cycle de vie d'une information, point de vue lié à l'expertise de l'interlocuteur (cadre réglementaire, juridique, culturel...)	Connaissances modélisées de l'interlocuteur
sémantique Établit les relations entre symboles élémentaires (p.ex. relation symbole / toponyme, légende)	Ensemble des traces situées du document, représentables et intervenant dans la mise en œuvre d'une intention de l'interlocuteur	Règles d'associations logiques
syntactique Assure la cohérence de la description par rapport à un référentiel	Formes et graphismes, chaînes de caractères, symboles	Règles grammaticales
topologique Organise les relations existantes entre chaque sous-ensemble	textures, topologie arborescente, listes chaînées d'objets	Règles d'associations formelles
morphologique Attribue des propriétés descriptives d'un sous-ensemble	Surface (raster), chain-code, polygone ou polygone (vecteur), invariants (raster), labels de primitives graphiques, caricatures (trans-reconstruction)	Descripteurs
inclusion Organise en sous-ensembles physiques des éléments répondant à des critères identiques	Masse connexe (raster), cluster (raster), segment (vecteur)	Information de segmentation
unaire Décrit l'espace de représentation des données au niveau physique du système	Pixel (raster), point (vecteur)...	Informations spatiales, spectrales, codage au niveau physique

6. Modélisation

L'activité autour du document, au sein d'une organisation, est un système dont la propriété la plus contraignante est la complexité. Comme tout système complexe, il

doit être abordé avec méthode, et pour ce faire, nous utiliserons les principes de la systémique tels que définis dans [BER73] et repris par [LEM94].

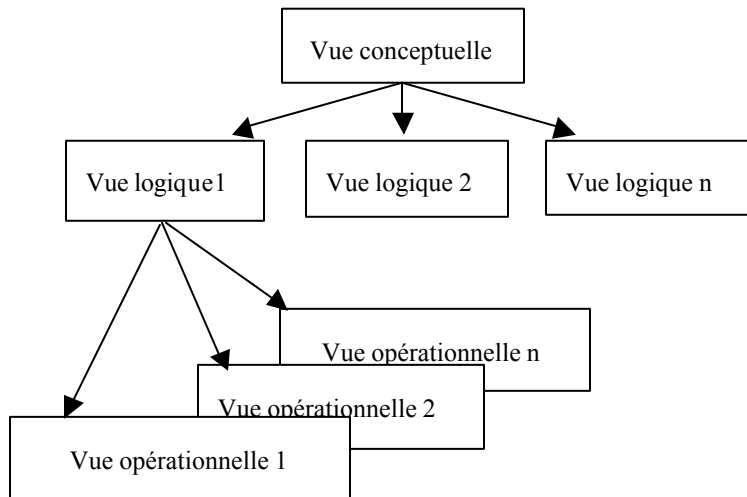
Modéliser un système complexe, c'est abstraire en préservant ce qui en fait la réalité, c'est-à-dire la complexité. Nous pouvons distinguer trois axes d'examen de l'activité autour du document : conceptuel, logique, et opérationnel.

Le **conceptuel**, c'est le fond essentiel du système, ici les fonctions fondamentales en ce qui concerne les activités, et les intentions portant sur l'information et la communication médiatisés par le document pour le contenu. Ces intentions et connaissances sont modélisées dans des ontologies, réseaux sémantiques, hiérarchies de concepts, réseaux mnésiques ou autres.

Le **point de vue logique** considère quant à lui les dispositions agencées dans le but de prendre en charge et réaliser les objectifs et contraintes définis par la description conceptuelle. Nous passons à ce niveau du quoi au comment. En ce qui concerne le document dans l'organisation, nous parlerons de scénarii d'intervention sur le flux d'information, et de structure logique de présentation de cette information.

Enfin, le **niveau opérationnel** représente les moyens techniques et les ressources permettant de réaliser les deux autres aspects : papier ou électronique, duplication ou unicité, logiciels de GED, etc.

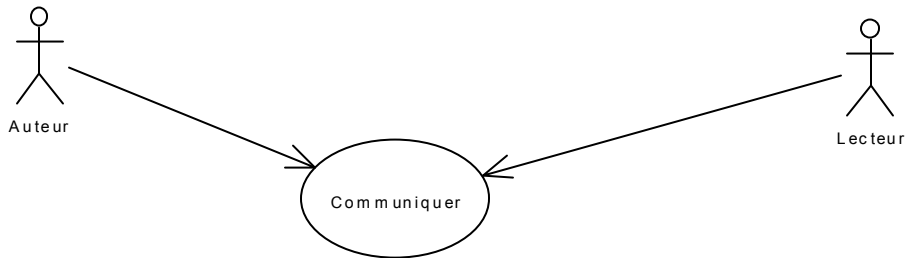
Ces trois niveaux sont bien sûr indissociables et en constante interaction, puisqu'il s'agit ni plus ni moins que de trois points de vue d'abstraction différents sur une seule et même réalité.



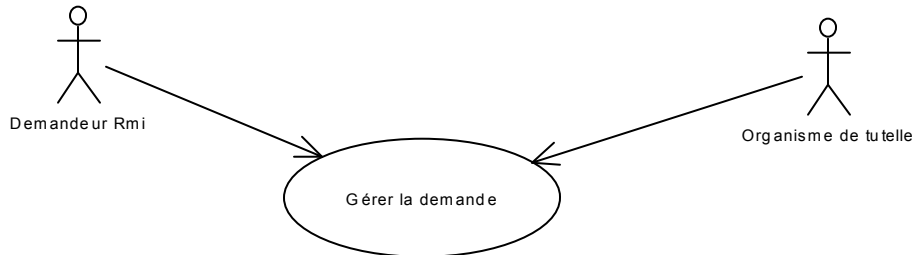
Modéliser dans notre AS permet de poser la question première : le document, pour quoi faire ? puis de considérer les possibles organisés permettant d'atteindre au mieux ce «pour quoi». Les choix opérés alors pourront se réaliser de multiples façons, les moyens matériels proposés étant nombreux et variés. En résumé, une seule vision conceptuelle, agenceable logiquement de diverses manières. Chaque choix logique est réalisable quant à lui à l'aide de divers moyens opérationnels.

Démarche :

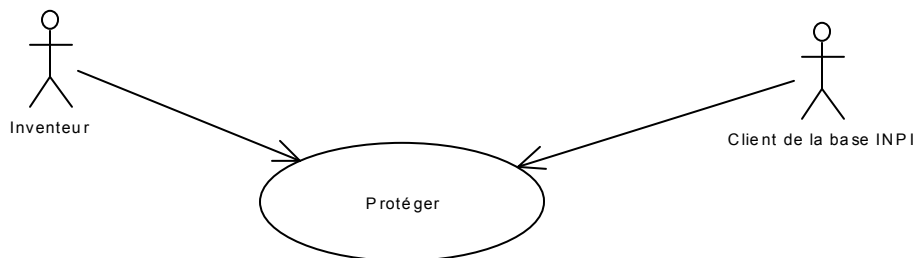
En Uml, la fonction essentielle (vue conceptuelle) de notre système est représentée par les cas d'utilisation suivants :



Pour le RMI :



Pour le brevet :



Les contraintes fonctionnelles restent vraies quelles que soient les dispositions prises pour les satisfaire. Notamment, on constate ici que la prise en charge par une organisation fortement structurée de la réalisation des objectifs conceptuels (téléologie) vient contraindre notablement le déroulement du processus, et la suite de l'analyse montrera qu'elle provoque un appauvrissement préjudiciable du contenu.

Tant que la communication entre auteur et lecteur s'opère directement, l'on met en prise directe les intentions et expertises de ces deux acteurs, qui peuvent par un jeu de dialogue, ajuster leurs compréhensions du contenu. Ce partage interpersonnel, envisageable conceptuellement devient irréalisable dès lors qu'intervient un tiers (ici l'organisme instructeur pour le Rmi, et l'INPI pour le brevet) chargé de recueillir le discours de l'auteur, et de le restituer sous une forme contrainte au lecteur.

C'est au niveau logique qu'interviennent ces tiers, et c'est leur survenue qui justifie la médiation par le document. Celui-ci a alors le rôle d'une matérialisation d'un contenu sémantique, interpersonnel et consensuel.

Ce consensus n'est obtenu que si l'on s'accorde sur l'identité et le rôle des acteurs, le sens du contenu (à l'aide d'un référentiel, e.g une ontologie ou autre), la forme imposée par le document, et les modes de transmission.

C'est à l'aide de modèles simples mais explicites et partageables que nous pourrions nous assurer de cet accord.

La modélisation d'un tel processus dans un environnement à forte composante logicielle demande une analyse particulière du système global. Cette analyse a pour objectif l'abstraction des seuls éléments modélisables dans ce contexte.

Du point de vue fonctionnel, une activité organisée peut-être vue comme étant une fonction admettant en entrée des flux d'informations, et générant en sortie des informations transformées (e.g des commandes de clients transformés en produits livrés et facturés, des étudiants à former, transformés en étudiants diplômés, des demandes de RMI transformées en allocations effectives, une invention transformée en protection industrielle, etc).

Le flux d'informations d'entrée traverse donc l'organisation considérée, et subit de sa part une succession de changements d'états.

Chaque fois que l'on constate un changement d'état du flux d'information, l'évènement ou l'activité l'ayant provoqué doit être identifié et décrit. C'est exactement sur ces transitions que viennent se greffer les activités et stratégies d'acteurs présentées au (§ 2).

En l'espèce, le document, (annoncé en introduction (§ 1) comme ensemble de données intermédiaires dans un processus de transformation), devient le support du flux trans-fonctionnel que nous chercherons à modéliser.

Les modèles proposés seront dans un premier temps spécifiques aux activités étudiées (RMI et brevet) puis élevés à un niveau générique. Nous emploierons pour ce faire le langage de représentation UML (Unified Modeling Language) normalisé par l'OMG (Object Management Group), et destiné à représenter les objets au sein d'un système, dans leurs états statiques et dynamiques.

Un examen de la pratique autour des demandes RMI peut être retracé à l'aide du diagramme d'état de la figure 3 :

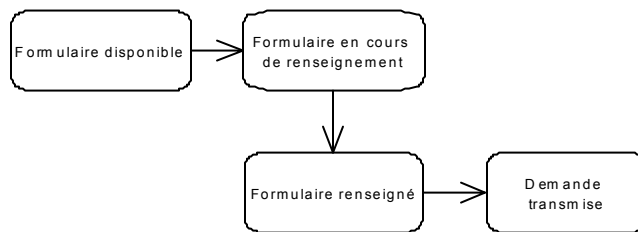


Figure 3

Le même exercice sur les documents brevet produit le diagramme d'état de la figure ci dessous.

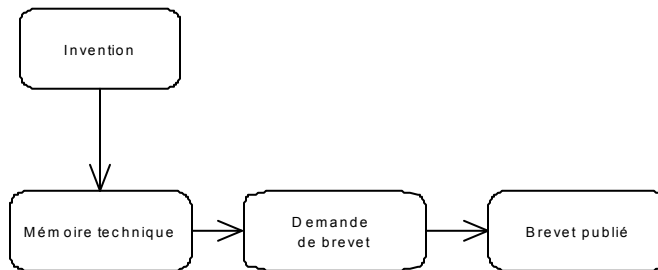


Figure 4

Il s'agit ici d'une représentation très superficielle destinée à exemplariser le propos. Il va de soi qu'une étude réelle irait à un niveau de détails et de granularité beaucoup plus avancé.

De ces diagrammes d'états, nous pouvons tirer les représentations d'activités suivantes pour le Rmi (figure 5) et pour le brevet (figure 6).

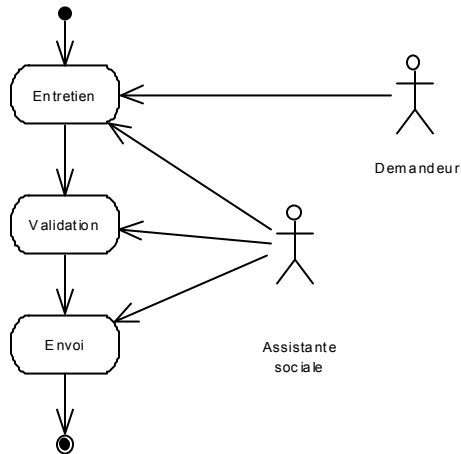


Figure 5

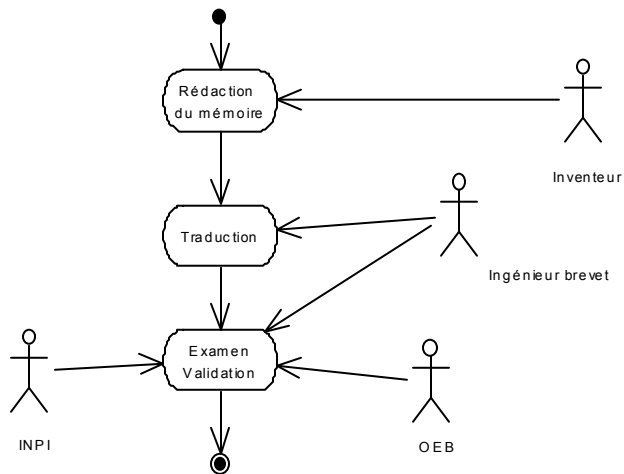
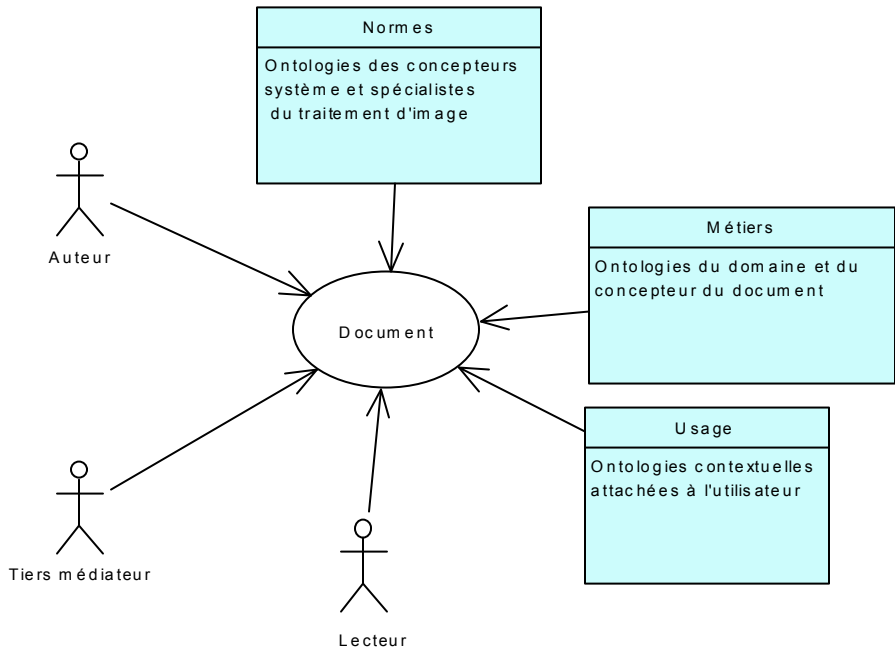


Figure 6

La question soulevée ensuite est celle des collaborations entre objets et acteurs permettant la réalisation de ces activités . Ces collaborations permettent entre autres de s'assurer d'un partage consensuel de sens sur les objets partagés par l'ensemble de la communauté concernée.



Les passages successifs « conceptuel->logique->opérationnel » apportent des contraintes de plus en plus lourdes venant dégrader les objectifs d'origine.

Considérer le document dans l'organisation comme étant la forme initiale et prévalente de l'information interpersonnelle revient à privilégier les contraintes d'ordre logique et technique sur le point de vue conceptuel.

A contrario, redonner au document le statut d'une présentation aboutie multimodale, s'appuyant sur un partage des connaissances et expertises au service des objectifs de l'organisation permet d'éviter l'écueil.

7. Conclusion

L'action spécifique "document et organisation" avait pour but initial d'étudier les processus d'échange d'informations au travers de documents dans un réseau représentant soit une même communauté d'intérêt, soit des communautés d'intérêts différentes.

Les discussions et confrontations scientifiques autour des deux terrains ont montré les limites de la modélisation en situation complexe. L'introduction de la notion d'interconnexion d'infosphères pourrait apporter des éléments de réponse en permettant l'instauration d'un dialogue réel homme – machine ou homme – homme ou homme – machine – homme qui autorise le partage des connaissances nécessaires à l'échange d'informations.

Les contraintes nées des organisations nous renseignent sur les protocoles liés aux échanges de documents, donc sur les codes nécessaires à la compréhension des informations.

Les méthodologies qui sont à construire dans le domaine des sciences pour l'ingénieur pour concevoir et mettre en œuvre les interfaces de dialogue des plateformes de traitement assisté des documents permettront également aux chercheurs des autres domaines scientifiques de mener les expériences permettant d'éprouver leurs théories.

8. Bibliographie

- [BER73] Von Bertalanffy L. ; General System Theory Ludwig London: Penguin, 1973.
- [DEN03] Dengel, A.R.(2003). Making document work: challenges for document understanding. Int. Conf. on Document Analysis and Recognition, pp. 1026–1035, Edinbourg, Scotland, 2003.
- [LEM94] Lemoigne JL. ; La théorie du système générale ; PUF 1994
- [NEW82] Newell, A.(1982). The knowledge level. *Artificial Intelligence*, Vol.19(N.2):pp. 87–127, 1982.
- [NON94] Nonak, I.(1994). Dynamic theory of organisation knowledge creation. *Organizational Science*, Vol.5(N.1):pp. 14–37, 1994.
- [PED03] Roger T Pédaque, Document : forme, signe et médium, les re-formulations du numérique