HAL
open science

# Le partage de documents scientifiques : problématique et questionnements

Françoise Armand, Pierre-Michel Riccio

## To cite this version:

Françoise Armand, Pierre-Michel Riccio. Le partage de documents scientifiques : problématique et questionnements. 2012. hal-00812707

**HAL Id: hal-00812707**

**https://hal.science/hal-00812707v1**

Preprint submitted on 12 Apr 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Le partage de documents scientifiques : problématique et questionnements

Françoise ARMAND
Documentaliste
Ecole Nationale Supérieure des Mines d'Alès

Pierre-Michel RICCIO
Maître de recherche
Centre de recherche LGI2P de l'Ecole Nationale Supérieure des Mines d'Alès

L'individu en situation est doté d'un système de pertinence composé de ses acquis (éducation, apprentissage, culture, …) et d'une vision du monde (Alfred Schutz, 1987). Cela signifie que deux individus, même assez proches, parce qu'ils n'ont pas exactement le même parcours de vie, possèdent nécessairement un système de pertinence, une vision du monde différents. Exprimé autrement : chaque individu perçoit la situation à travers un filtre personnel, il n'y a pas une mais plusieurs réalités.

Dans une situation d'interaction, l'individu, doté d'une rationalité limitée (Herbert Simon, 1976), va mécaniquement chercher à apparier phénomènes observés et connaissances acquises pour prendre une décision (Gary Klein, 1998). Ce mécanisme, mis en évidence dans les situations d'urgence, fonctionne de la même façon dans les situations d'interaction.

Les différences d'interprétation peuvent alors conduire les individus à suggérer ou prendre des décisions décalées, contradictoires, opposées, … ce qui va nuire à l'efficacité du collectif dans l'action. Dans un contexte de concurrence exacerbée, la performance d'un collectif scientifique ne peut être réduite au fait de rassembler les informations « utiles » dans des réservoirs de données. Pour travailler ensemble de façon plus efficace, il est nécessaire de favoriser la recherche d'une compréhension partagée en faisant converger – au moins en partie – le système de pertinence et la vision du monde des individus impliqués.

## 1 – Un rapide panorama de la situation

Le développement très rapide des réseaux numériques et l'intensification de la concurrence internationale font que le partage d'informations et de connaissances est devenu un enjeu majeur pour les scientifiques. Les supports de communications – actes de colloques et de journées d'étude, revues papier et électronique, ouvrages – se multiplient et il existe une pression constante pour inciter les scientifiques à publier leurs travaux sur les meilleurs supports.

L'ensemble fait que nous pouvons constater aujourd'hui un foisonnement d'articles scientifiques : en accès libre sur les réseaux numériques, en accès libre dans des réservoirs de données publics comme HAL ou arXiv, accessibles via des abonnements chez les éditeurs, ou librement partagés via des plateformes de travail collaboratif locales.

L'organisation est toutefois très différente selon les spécialités scientifiques. Sous l'impulsion d'autorités comme l'U.S. National Library of Medecine, des spécialités comme la biologie et la médecine disposent aujourd'hui d'outils performants comme PubMed, plébiscité par les scientifiques. PubMed propose un ensemble assez exhaustif de résumés d'articles scientifiques accompagnés de liens facilitant l'accès aux documents sur le site des éditeurs (via des abonnements généralement pris en charge par les institutions). Mais, un grand nombre de spécialités ne sont pas aussi bien organisées.

Si le dépôt d'articles et de résumés dans des archives ouvertes comme HAL est recommandé en France, force est de constater que la démarche pour le moment n'est pas généralisée et que les documents déposés ne représentent qu'une partie limitée de la production scientifique française.

Toutefois, même si la démarche qui consiste à rassembler les informations potentiellement « utiles » dans des réservoirs de données est indispensable, elle semble ne pas être suffisante. En effet, plusieurs questions surgissent : Quelles informations vont permettre à un scientifique d'avancer plus facilement dans ses travaux ? Que cherche un scientifique dans les réservoirs de données ? Comment évaluer la pertinence, l'utilisabilité des informations stockées dans ces réservoirs ? Comment indexer les documents pour qu'un individu les trouve plus facilement ? Quels dispositifs imaginer pour améliorer la performance des collectifs scientifiques ?

## 2 – Des pistes de travail

Les études montrent que les utilisateurs déploient des stratégies différentes pour chercher de l'information (Nicole Boubée, 2010). Les experts, dans notre cas les scientifiques expérimentés, auraient tendance : à formuler des requêtes appropriées, à s'aventurer plus en profondeur dans les environnements

documentaires, à privilégier l'accès direct aux ressources de préférence à l'utilisation de moteurs de recherche, et à examiner un nombre réduit de documents ciblés. Les novices, dans notre cas les doctorants et scientifiques débutants, consacreraient peu de temps à la planification d'une stratégie de recherche d'informations et privilégieraient le butinage. Pour certains spécialistes la typologie des utilisateurs est plus complexe : il convient de prendre en compte l'expérience dans le domaine, mais aussi l'expérience dans la recherche d'informations (Christoph Höelscher, 2000). Dans nos pratiques, nous avons constaté que les scientifiques expérimentés privilégient la recherche d'informations par auteurs, alors que les scientifiques moins expérimentés ont tendance à privilégier une recherche par thèmes / mots clés.

Alors que les motivations sont multiples, la concurrence exacerbée, et compte tenu de la multiplicité des ressources accessibles sur les réseaux numériques, comment imaginer un ou des dispositifs adaptés pour faciliter la recherche et / ou le partage des informations scientifiques, la construction collective des connaissances ?

En s'appuyant sur les expériences réussies dans le domaine de la biologie et de la médecine, et si l'on considère qu'il est aujourd'hui possible de rassembler en un même lieu les informations importantes, quitte si nécessaire à privilégier un rapide descriptif des documents accompagnés de liens pérennes pointant sur les documents originaux (dans le cas où pour des raisons techniques ou juridiques ces derniers ne peuvent être déplacés), plusieurs points sont à considérer.

La façon dont les documents sont indexés, classés, hiérarchisés, référencés est essentielle. La question n'est pas nouvelle puisque les premières applications de gestion des documents ont été basées sur la classification Dewey de 1876 (Melvil Dewey, 2004). Même si les techniques ont ensuite évolué (Paul Otlet et Henri La Fontaine, 2001), l'idée consiste toujours à décrire l'ensemble des documents à l'aide de descripteurs ou mots-clés en s'appuyant sur des référentiels ou thésaurus. La nouvelle norme ISO 25964-1 témoigne de cette continuelle volonté de codifier les outils d'indexation afin de fluidifier la recherche documentaire et l'interopérabilité des données. Même si de nombreux spécialistes s'interrogent sur l'utilité de créer de nouveaux thésaurus, nous constatons que la construction d'une taxinomie des thèmes est nécessaire car elle correspond à un usage, en l'occurrence ordonner, classer, tagger des documents / notices pour pouvoir les retrouver facilement en maximisant les réponses utiles et minimisant le bruit de fond.

De nombreux utilisateurs préfèrent se servir d'un moteur de recherche qui, en s'appuyant sur un thésaurus et l'ensemble des mots présents dans les documents et leurs annexes, va leur permettre en réponse à une requête

d'obtenir facilement une liste de documents. L'usage des moteurs de recherche s'est progressivement généralisé dans nos pratiques au quotidien. Toutefois, il faut bien reconnaître que, même pour un expert capable de décrire avec beaucoup de précision sa requête pour faciliter la discrimination, le volume considérable d'informations rapportées par une requête généraliste conduit *de facto* à faire confiance aux choix réalisés par le moteur et les concepteurs de ce moteur dans la construction d'un ordre de pertinence décroissant, et ceci n'est pas nécessairement une garantie de qualité.

Les travaux dans le domaine du web sémantique ont fait ces dernières années des progrès considérables. L'émergence et l'utilisation d'ontologies – spécification explicite (formelle) d'une conceptualisation (partagée) (Tom Gruber, 1993) – permet aujourd'hui d'imaginer des moteurs de recherches capables de s'appuyer sur des concepts partagés (et non exclusivement sur des mots) pour trouver les documents et ordonnancer de façon plus pertinente les listes de documents en réponse à une requête. Par ailleurs, de nouvelles approches comme celles des folksonomies sont apparues. L'idée, de façon simplifiée, consiste à donner la possibilité à un ensemble d'utilisateurs de commenter les documents et intégrer tout ou partie de ces commentaires dans le modèle de connaissances en construction.

Comme nous le précisions dans l'introduction de cet article, notre parti pris est de considérer en premier lieu l'action de groupes de scientifiques engagés dans une action commune.

Aussi, comment construire ces modèles de connaissances ? Faut-il imaginer – à l'image de ce que propose IEML (Pierre Levy, 2006) – un modèle universel pour harmoniser, clarifier, au niveau international le point de vue de l'ensemble des spécialistes d'un domaine, ou bien privilégier des conceptualisations partagées locales satisfaisantes pour un nombre limité de scientifiques impliqués dans des actions communes ?

## 3 - DES CRITERES A PRIVILEGIER

Dans les années 60, le philosophe des sciences Thomas Kuhn nous a montré que les scientifiques partagent un ensemble de connaissances appelé aussi paradigme jusqu'au moment où émerge une meilleure approche ou théorie qui provoque un changement de paradigme et une reconfiguration du périmètre des travaux (Thomas Kuhn, 1972).

Dans les années 70, le philosophe français Michel Foucault a proposé de revisiter la notion de dispositif : « un ensemble résolument hétérogène, comportant des discours, des institutions, des aménagements architecturaux, des décisions réglementaires, des lois, des mesures administratives, des énoncés

scientifiques, des propositions philosophiques, morales, philanthropiques, bref : du dit, aussi bien que du non-dit » (Michel Foucault, 1994).

Pour faciliter la création d'un dispositif au sens de Foucault, qui facilite le partage d'informations et de connaissances au sens de Kuhn, la construction d'ontologies – reflets d'un sens commun partagé par un groupe d'individus engagés dans une action commune – au sens de Gruber, nous avons choisi de porter notre attention sur trois composantes ou « critères » à privilégier : le point critique de satisfaction, la réutilisation / transposition des ontologies, et le paradoxe du scientifique.

- Le point critique de satisfaction : Pour pallier l'antagonisme d'un double processus, évolution rapide des thématiques versus stabilisation des outils pour partager des connaissances, il faut accepter l'idée d'un compromis : l'outil n'est pas omniscient mais doit en contrepartie être stable et faciliter une utilisation intuitive. Même imparfait, il peut alors soutenir de façon efficace le processus de construction d'un modèle partagé des connaissances, processus complexe et pas toujours rationnel.
- La réutilisation des ontologies : Dans la construction d'ontologies (Bruno Bachimont, 2000), le critère « réutilisation » est, à notre avis, essentiel. En effet, sous le prétexte légitime de créer un outil reflétant une réalité de terrain, les scientifiques ont tendance à construire de façon systématique de nouvelles ontologies. Notre idée est de rompre avec le modèle de connaissances de type « haute couture » pour privilégier des modèles de connaissances « prêt à porter » plus facilement réutilisables.
- Le paradoxe du scientifique : Le scientifique se trouve au centre d'un champ de forces contradictoires : besoin de partager, et nécessité de conduire un travail personnel pour répondre aux injonctions de l'évaluation (Bruno Latour, 2001). Aussi, notre idée est de prendre en compte dans la construction du dispositif de partage d'informations et de connaissances, l'indispensable négociation entre acteurs porteurs d'enjeux personnels et collectifs, et qui doivent travailler ensemble (Véronique Richard, 2009).

## 4 - UN DEBUT DE REPONSE

De nombreux travaux en sciences et technologies de l'information et de la communication sont dédiés à la construction de modèles d'information et de connaissances : ontologies, folksonomies, fouille de données, … De la même façon, de nombreux travaux en sciences humaines et sociales s'intéressent aux usages et pratiques des technologies numériques : profil des utilisateurs, enjeux et motivations, normes et contraintes,… Cependant, peu de travaux de

recherche sont centrés sur la construction collective de modèles de connaissances souples et stables, transposables, insérés dans un dispositif qui favorise – dans un processus de négociation – l'efficacité d'une action commune comme la valorisation des actions / enjeux individuels.

C'est l'objectif du travail de recherche que nous proposons, centré dans une convergence disciplinaire sur la création d'une démarche de construction collective de dispositifs de partage d'informations et de connaissances dans un contexte scientifique. Les enjeux en termes d'impact sont importants, et ce travail se situe dans la continuité de plusieurs projets de recherche comme le projet français DaFOE ou encore le projet européen NeOn (www.neon-project.org).

Le projet DaFOE est un projet ANR piloté par Jean Charlet dans lequel plusieurs équipes françaises sont impliquées. La construction collective de modèles de connaissances est un processus très complexe, généralement décomposé en quatre étapes (constitution d'un corpus de documents, analyse linguistique du corpus, conceptualisation, opérationnalisation de l'ontologie). Il existe pour chacune de ces étapes de nombreux outils, mais « aucune procédure généralement acceptée » et « aucun ensemble cohérent d'outils support » (Jean Charlet, 2008). Aussi, l'ambition de DaFOE est de proposer une architecture générale, une plateforme modulaire, destinée à faciliter la construction de modèles de connaissances, d'ontologies dans le respect des pratiques habituelles (méthodes différentes et modélisation dynamique).

Si la démarche retenue pour le projet DaFOE consiste avant tout à penser l'interopérabilité entre les différents outils, il convient maintenant (au terme du projet) d'évaluer son usage en situation. Au-delà de l'évaluation intrinsèque de la plateforme, qui va permettre comme le précisent les auteurs d'évaluer la persistance, la traçabilité et le dimensionnement des modèles de connaissances, il nous semble – compte-tenu des critères que nous avons choisi de privilégier – nécessaire d'évaluer cette proposition d'architecture au regard : du point critique de satisfaction, de la réutilisation / transposition des ontologies, et du paradoxe du scientifique.

## CONCLUSION

Notre objectif – fondé sur les travaux que nous conduisons depuis plusieurs années sur le thème de l'intelligence collective dans une convergence disciplinaire : à la rencontre des technologies numériques et des sciences humaines (Monique Commandré et Pierre-Michel Riccio, 2010) – est d'imaginer un dispositif facilitant le partage d'informations et de connaissances scientifiques.

L'originalité de ce travail de recherche n'est pas d'initier une nouvelle approche *ex nihilo*, mais, en nous appuyant sur des travaux existants, d'appréhender les usages et pratiques pour aménager les dispositifs dans l'action et ce dans une perspective interdisciplinaire. Nous nous plaçons de ce fait dans une démarche volontariste de compromis : cherchant du point de vue des SHS à privilégier la facilité de prise en main d'un outil de travail et du point de vue des STIC à créer un outil souple et robuste. L'ensemble imaginé pour renforcer l'efficacité du collectif dans l'action doit apporter un premier niveau de réponse sur les critères que nous avons choisis de privilégier : construire des modèles de connaissances partagés, transposables, adaptés aux spécificités des scientifiques en situation.

**Bibliographie**

BACHIMONT B., Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances, in Ingénierie des connaissances : évolutions récentes et nouveaux défis, éditeurs CHARLET J. ZACKLAD M., KASSEL G. et BOURIGAULT D., Eyrolles, Paris, 2000.

BOUBEE N. et TRICOT A., Qu'est-ce que rechercher de l'information, Presses de l'ENSIB, 2010.

CHARLET J., SZULMAN S. and PIERRA G., DAFOE : A multimodel and Multimethod Platform for Building Domain Ontologies, JFO 2008, 1-2 décembre 2008, Lyon France, ACM, 2008.

COMMANDRE M. et RICCIO P.M., A travers les ontologies : vers de nouveaux agencements pour l'organisation et l'accès à l'information scientifique, Actes du colloque Informations et organisations : nouvelles stratégies structures et fonctions (COSSI'2010), Université de Moncton, Campus de Shippagan, NB Canada, 16 et 17 juin 2010.

DEWEY M., A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library, Gutenberg Project, http://www.gutenberg.org/etext/12513, 2004.

FOUCAULT M., Le jeu de Michel Foucault, Dits et écrits, T. III, Gallimard, Paris, 1994, 298-329.

GRUBER T.R., A translation approach to portable ontology specifications, Knowledge acquisition, 5 (2), 1993, 199-220.

HOELSCHER C. and STRUBE G., Web search behavior of Internet experts and newbies, Computer Networks, 33 (1-6), 2000, 337-346.

KLEIN G., Sources of Power How People Make Decisions, MIT Press, 1998.

KUHN T.S., La structure des révolutions scientifiques, Flammarion, Paris, 1972.

LATOUR B., Le métier de chercheur, regard d'un anthropologue, INRA Editions, Paris, 2001.

LEVY P., IEML : finalités et structure fondamentale, in Intelligence Collective : Rencontres 2006, Presses des Mines de Paris, 2006, 117-136.

OTLET P. et LA FONTAINE H., Classification décimale universelle : édition abrégée (7ème éd), Editions du Céfal, Liège, 2001, 292 p.

RICHARD V., Pour une éthique relationnelle dans les organisations contemporaines ? In Management des Technologies Organisationnelles (MTO'2009), Presses des mines, Paris, 2009.

SCHUTZ A., Le chercheur et le quotidien, Méridiens Klincksieck, Paris, 1987.

SIMON H., Administrative behavior: A study of decision-making processes in administrative organizations (3rd ed.), Free Press, New-York, NY, 1976.