

Edition électronique des constructions scientifiques et auto-archivage des bases de données

Valentine Roux (UMR 7055, Nanterre)

Nous connaissons actuellement une crise au sein des publications scientifiques, souvent évoquée par des chercheurs provenant de disciplines diverses¹:

- d'une part, il n'est plus possible, à l'échelle individuelle, de consommer le nombre croissant d'écrits touchant à des domaines de recherche spécifiés ; ce problème biodémographique du temps de lecture pénalise fortement le cumul des connaissances, de même que les échanges interdisciplinaires qui nécessitent, de fait, l'élargissement des lectures.
- d'autre part, les coûts d'édition imposent de publier les résultats sans les données, privant ainsi le lecteur de l'appareil critique nécessaire à l'évaluation du fondement des constructions avancées.
- enfin, les publications scientifiques n'étant plus le lieu de publier les bases de données, le chercheur se trouve devant un problème de cumul d'archives vivantes que les campagnes de numérisation sont loin d'avoir résolu (problème d'indexation et d'organisation des archives).

Dans le domaine de l'ingénierie des connaissances, ce problème de publications est bien connu sous le vocable « Cognitive Overflow Syndrom »². Celui-ci décrit la surcharge du travail intellectuel liée à l'abondance de l'information à traiter qui ne fait que s'accroître avec l'accès généralisé à des sources documentaires électroniques. Cette multiplicité de l'information se traduit par le fait que le travailleur intellectuel ne parvient plus à produire du sens à partir de la multitude d'informations auxquelles il peut accéder.

En SHS, les écrits linéaires posent, en outre, le problème de la visibilité des pratiques scientifiques, et par suite, celui du transfert des connaissances tant à l'endroit des scientifiques, que des étudiants ou des non scientifiques³.

Contexte scientifique

En réponse à cette surcharge d'informations, ainsi que cette absence de lisibilité, deux

1 On pourra citer notamment Pierre Joliot (titulaire au Collège de France de la chaire de bio-énergétique cellulaire) : « Il faut tout d'abord souligner que les chercheurs n'ont pas attendu l'arrivée des réseaux de communication pour être submergés par un flot de publications dépassant leurs capacités d'assimilation. En se limitant aux articles publiés dans les journaux scientifiques, il est depuis longtemps impossible d'assimiler toute l'information disponible dans son propre domaine de recherche. » (Joliot, 2001 :88). Voir également Gardin, J.C. 1998. *Cognitive issues and problems of publication in archaeology*. In Hensel, W., Tabaczynski, S. et Urbanczyk, P. (Eds.) *Theory and practice of archaeological research*. Warszawa : Institute of Archaeology and Ethnology, Committee of pre- and protohistoric sciences, Polish Academy of Sciences, p.65-113. Gardin, J.C. 1999. *La suproduction des publications en sciences humaines: ses rapports avec la question du mélange des genres*. Document de travail du séminaire « Le Modèle et le Récit », consultable à la bibliothèque de la MSH.

2 Lahlou, S. (200/1) La cognition au travail et ses outils : débordement, révolution, distribution, *Intellectica*, 30, 7-17.

3 (sur ce sujet, voir P. Geslin, D. Magda, P. Salembier, C. Albaladejo *La transférabilité des connaissances en question(s)*. Paris : Editions du CNRS)

positions ont été prises.

Dans le domaine des NTIC, des recherches ont été engagées qui visent à élaborer des outils d'extraction des données et de rationalisation des stratégies d'accès à l'information. Ces outils font essentiellement référence aux outils documentaires, aux résumés automatiques, aux fonctions de filtrage et à l'organisation sémantique des informations ou des données (cf par exemple les travaux récents sur le web sémantique).

En SHS, des recherches sont engagées sur la modélisation des textes scientifiques. Elles font référence au débat épistémologique qui distingue entre le modèle et le récit. Deux ouvrages récents ont publié ces débats⁴.

Parmi les différentes solutions proposées, J.-C. Gardin a développé un programme logiciste qui propose de condenser les textes scientifiques en distinguant les différents noyaux cognitifs qui en organisent les constructions⁵. Ces différents noyaux cognitifs distinguent la base de faits, les conclusions ou hypothèses finales et les opérations logico-discursives permettant de passer des premières aux secondes. Gardin utilise à dessein le terme de condensation dans le sens où la modélisation doit permettre de réduire le volume des textes sans en perdre la substance.

Ce programme logiciste est toujours l'objet de discussions nourries, ceci à une échelle internationale et transdisciplinaire. Toutefois, il a été peu mis en œuvre, l'une des raisons avancées étant que la lecture des diagrammes logicistes sur papier était mal aisée. L'arrivée des NTIC a permis de reconsidérer la présentation de ces schématisations et de proposer une écriture multimédia qui assure à la fois la manipulation aisée des constructions scientifiques et l'accès à l'ensemble des données qui les fondent.

Ainsi une expérimentation consistant à modéliser sur multimédia l'ensemble d'un ouvrage⁶ a été menée.

Le modèle éditorial de ce cédérom a été conçu pour que le chercheur puisse consulter rapidement les grandes étapes des constructions scientifiques, les principaux résultats obtenus à chaque étape, les arguments relatifs à chaque résultat, les séries de données mobilisées par les arguments. Cette modélisation a été appliquée aux études de corpus archéologiques et aux référentiels destinés à leur interprétation. Ce sont ainsi des textes fort divers qui ont été modélisés, tant du point de vue des disciplines dont ils sont issus (archéologie, psychologie, économie, sciences des matériaux), que des données traitées (objets, chiffres, schémas qualitatifs, films, imagerie microscopique...).

Quatre écrans successifs organisent la consultation ; chacun d'entre eux livre un niveau de lecture. Les quatre niveaux de lecture sont indiqués au tableau 1.

Dans le cadre d'une édition multimédia, ces quatre niveaux d'information sont accessibles de manière dynamique fonctionnant à la manière de zooms successifs. Ils permettent la consultation rapide des résultats, de leur fondement et des données mobilisées. Le support numérique permet, en outre, l'indexation raisonnée de la totalité des séries de données, assurant ainsi un archivage des faits scientifiques ayant servi à la construction d'un résultat.

4 Grenier, J.Y., Grignon, C. et Menger, P.M. (Eds.) 2001. *Le Modèle et le Récit*. Paris : Editions de la MSH. Berthelot, J.-M. (sous la dir.) 2000. *L'épistémologie des sciences sociales*. Paris : Presses Universitaires de France.

5 Gardin, J.C. 1979. *Une archéologie théorique*. Paris : Hachette. Gardin, J.C. 1991. *Le calcul et la raison. Essais sur la formalisation du discours savant*. Paris : Editions de l'EHESS.

6 « Cornaline de l'Inde. Des pratiques techniques de Cambay aux techno-systèmes de l'Indus », sous la dir. V. Roux, éditions de la MSH, Paris, 2000

1 ^{er} écran	Énoncé du plan d'étude ou encore des principales opérations qui structurent la construction scientifique.
2 ^{ème} écran	Énoncé des propositions développées à chaque étape du plan. Cet écran est celui de la lecture rapide des principaux résultats.
3 ^{ème} écran	Énoncé des arguments qui étayent chaque proposition. La pertinence des arguments, pour étayer une proposition, est expliquée en langue naturelle dans un commentaire.
4 ^{ème} écran	Présentation des bases de données mobilisées par les arguments. Les données peuvent correspondre à des documents iconographiques très divers (graphiques, films, photos, animations...). Elles sont décrites en langue naturelle dans un champ réservé à cet effet.

Tableau 1. Contenu des 4 écrans interactifs correspondant aux 4 niveaux de consultation d'un article scientifique.

L'outil d'édition développé dans de CD-Rom "Cornaline de l'Inde" se caractérise ainsi de la manière suivante:

³⁵₁₇ il s'agit avant tout d'un **outil de lecture rapide** qui permet de consulter rapidement une construction scientifique et les données qu'elle mobilise. Tout se fait en 4 "clics", chaque clic est informatif. Un tel outil assure la visibilité de nos pratiques scientifiques et, par conséquent, le transfert des connaissances élaborées tant au sein de notre propre communauté qu'auprès d'autres disciplines, point particulièrement crucial à une époque où les projets interdisciplinaires nécessitent des dialogues méthodologiques parfois difficiles à mettre en œuvre.

³⁵₁₇ il s'agit d'un **outil d'auto-archivage** de données scientifiques. L'organisation des bases de données est relative aux constructions scientifiques. On peut considérer que leur archivage est achevé au terme de la publication. L'avantage de bases de données constituées relativement à des constructions scientifiques réside dans le système d'interrogation qui est guidé par les opérations à l'origine de leur organisation.

³⁵₁₇ enfin, il s'agit d'un **outil d'édition permettant de diminuer considérablement les coûts d'édition**. La saisie se fait de manière naturelle à l'écran par le chercheur. La relecture est facilitée par le volume restreint de texte. La publication des données ne présente pas de contraintes de volumes. La diminution des coûts est également patente en ce qui concerne la traduction des textes en plusieurs langues.

Publications

ROUX, V. sous presse. Formalisation des publications scientifiques sur support électronique. Un cas d'application à l'ouvrage archéologique "Cornaline de l'Inde". In A. IACOVELLA (ed.) *Sémantique et archéologie. Aspects expérimentaux. Renouvellements méthodologiques dans les bibliothèques numériques et les publications scientifiques*. Actes de la Table-ronde interdisciplinaire organisée par l'École française d'Athènes, Athènes, 18-19 novembre 2000, Bulletin de Correspondance Hellénique suppléments.

ROUX, V. sous presse. Modélisation des constructions scientifiques sur multimédia et transfert des connaissances. In P. GESLIN, D. MAGDA, P. SALEMBIER, C. ALBALADEJO *La transférabilité des connaissances en question(s)*. Paris : Editions du CNRS.

ROUX, V. et BLASCO, Ph. (conception et réalisation) 2000. *Cornaline de l'Inde*. CD-Rom bilingue (français, anglais). Paris : Editions de la MSH, 2000. (vendu avec l'ouvrage *Cornaline de l'Inde. Des pratiques techniques de Cambay aux techno-systèmes de l'Indus*, V. Roux (dir)).