



Ghislaine Chartron, Benoît Epron et Annaïg Mahé (dir.)

Pratiques documentaires numériques à l'université

Presses de l'enssib

Pratiques informationnelles en physique des particules

Anne Gentil-Beccot

DOI : 10.4000/books.pressesensib.1174

Éditeur : Presses de l'enssib

Lieu d'édition : Presses de l'enssib

Année d'édition : 2012

Date de mise en ligne : 20 juillet 2017

Collection : Papiers

ISBN électronique : 9782375460511



<http://books.openedition.org>

Référence électronique

GENTIL-BECCOT, Anne. *Pratiques informationnelles en physique des particules* In : *Pratiques documentaires numériques à l'université* [en ligne]. Villeurbanne : Presses de l'enssib, 2012 (généré le 01 février 2021). Disponible sur Internet : <<http://books.openedition.org/pressesensib/1174>>. ISBN : 9782375460511. DOI : <https://doi.org/10.4000/books.pressesensib.1174>.

PRATIQUES INFORMATIONNELLES EN PHYSIQUE DES PARTICULES

La physique des hautes énergies est une discipline qui vise à découvrir les constituants de la matière et comprendre leurs interactions. Elle compte environ 30 000 chercheurs dans différents instituts de recherche et universités répartis dans le monde et constitue une petite communauté scientifique qui possède une culture internationale et collaborative très marquée. D'autres particularités propres à cette communauté l'ont conduite à développer des pratiques informationnelles spécifiques. Elle a en effet adopté, dès les premières heures de la discipline, un système de communication scientifique unique, basé en grande partie sur la prépublication. Ces pratiques ont évolué avec le temps mais, aujourd'hui encore, les prépublications et la littérature grise en général restent une indispensable ressource pour cette discipline, elle est même devenue un moteur qui entraîne de nombreuses évolutions dans le paysage informationnel.

Dans le même temps, la publication dans les revues scientifiques demeure essentielle pour la communauté scientifique, la plupart des prépublications produites étant finalement publiées. En outre, les technologies de l'information deviennent de plus en plus complexes. En plus des outils développés au sein de la communauté, les scientifiques ont la possibilité d'utiliser de nombreux produits comme les bases de données commerciales ou des moteurs de recherche tels que Google ou Google Scholar ; les utilisateurs peuvent désormais accéder, facilement ou non, à d'énormes quantités d'informations. Par ailleurs, les technologies sont plus innovantes chaque jour et de nouvelles interactions avec les utilisateurs se créent. Dans un tel contexte, il est intéressant d'analyser les usages actuels en physique des particules en termes d'information scientifique et de comprendre les évolutions en cours.

Nous commencerons par une description de la culture de la prépublication en physique des particules et du développement des outils adaptés à ces ressources spécifiques, puis nous montrerons quels sont les bénéfices de ce schéma de communication pour la communauté. Enfin, nous

montrons comment elle adapte en permanence ses systèmes d'information en fonction des besoins des utilisateurs et des technologies nouvelles.

LA DIFFUSION DES PRÉPUBLICATIONS, UNE HABITUDE ANCRÉE DEPUIS DES DÉCENNIES

+++++

Dans les années 1950, le délai entre la soumission d'un article scientifique et sa mise à disposition pour le lecteur était inacceptable pour les physiciens des particules. La communauté était déjà composée de deux groupes pour qui la communication – au sein du groupe et avec l'extérieur – était essentielle : durant les premières heures de la discipline, les physiciens expérimentalistes travaillant sur les accélérateurs de particules d'une énergie sans cesse accrue observaient régulièrement de nouvelles découvertes et les théoriciens interprétant ces résultats amélioraient leurs théories et suggéraient de nouvelles expériences. Il était tout simplement hors de question d'accepter des mois de délai dans la diffusion des résultats de recherche – le temps moyen entre la naissance d'une idée et les premiers résultats expérimentaux était de quelques mois seulement –, les prépublications étaient donc la solution idéale.

Pendant des décennies, des physiciens théoriciens et expérimentalistes, désireux de diffuser rapidement leurs résultats de recherche, ont envoyé par courrier postal à tous les grands instituts de recherche en physique des particules des copies de leurs manuscrits tout en les soumettant en parallèle à des revues [Goldschmidt-Clermont, 2002 ; Heuer *et al.*, 2003 ; Aymar, 2009]. Les instituts soutenaient alors financièrement la diffusion des résultats scientifiques de leurs chercheurs, ce qui impliquait des coûts relativement élevés. Les bibliothèques de ces mêmes instituts consacraient également des ressources à l'indexation de toutes ces prépublications, les rendant accessibles aux chercheurs de l'organisation. Cette forme de libre accès assurait une diffusion massive et rapide des résultats scientifiques. Il est à noter que les chercheurs des institutions riches étaient forcément favorisés par ce système. Celles-ci pouvaient en effet payer pour l'envoi massif et étaient, par ailleurs, beaucoup plus susceptibles de recevoir des copies de prépublications d'autres scientifiques en quête de reconnaissance et de visibilité. Les institutions plus petites et plus pauvres avaient donc moins de chance de diffuser leurs résultats et d'être informées des recherches d'autres scientifiques.

Avec l'utilisation croissante d'Internet, le processus s'est poursuivi par voie électronique, ce qui a permis de rendre le système plus équitable.

Avec le développement des technologies, la communauté a créé ses propres outils pour gérer la littérature grise : en 1991, avant même que le Web ait été inventé, Paul Ginsparg, à Los Alamos, a lancé arXiv [arXiv, 2010], le premier serveur de prépublications en physique. Ce nouvel outil a assuré la transition entre une culture du papier obsolète et un nouveau processus électronique, en offrant à tous les scientifiques un moyen facile, gratuit et libre de diffuser leurs prépublications et en supprimant la barrière due au coût de diffusion par la poste des prépublications.

Avec plus de 500 000 articles, arXiv s'est aujourd'hui développé au-delà de la physique des hautes énergies, devenant ainsi le serveur majeur pour de nombreuses autres disciplines [Ginsparg, 1994].

SPIRES [SPIRES, 2010], le premier catalogue électronique de littérature grise [Addis, 1994 ; Kreitz et Brooks, 2003], est né à Stanford Linear Accelerator Center (SLAC), laboratoire à Stanford en Californie, en 1974, et a été développé en collaboration avec DESY, à Hambourg en Allemagne, et Fermilab, à Chicago. Ce catalogue contient à l'heure actuelle des métadonnées d'environ 760 000 articles dans le domaine.

Avec la création du Web, SPIRES est devenu le premier serveur Web [Kunz *et al.*] aux États-Unis en 1992, et des liens ont été créés vers arXiv pour le texte intégral, les prépublications sont donc enfin disponibles sur la toile, avec une accessibilité accrue grâce à l'indexation détaillée incluant les références aux versions publiées.

Aujourd'hui, la communauté produit environ 5 000 articles publiés par an. La grande majorité de ces articles sont publiés dans seulement six revues à comité de lecture [Mele *et al.*, 2006]. Dans la figure 1, on voit que 90 à 100 % des articles publiés dans ces six revues sont également soumis sur arXiv : nous voyons également que cette situation est stable et dure depuis dix ans déjà. En outre, de nombreux scientifiques soumettent de nouveau sur arXiv une version révisée de leur prépublication correspondant à la version finale publiée et incluant les corrections apportées au cours du processus de publication.

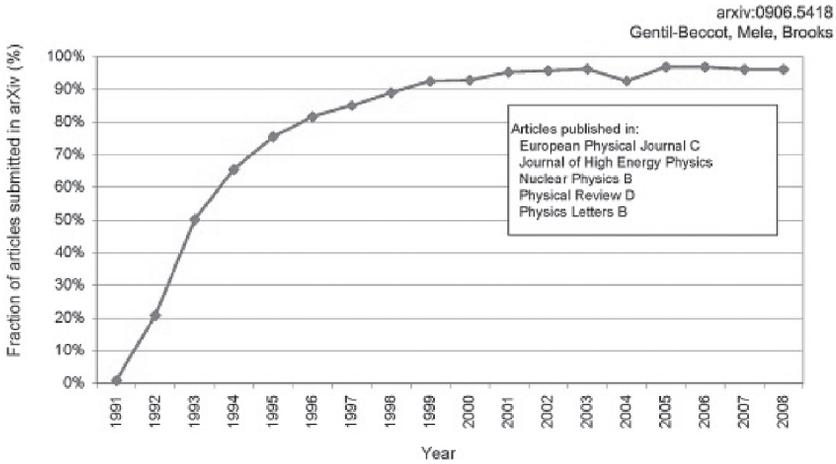


Figure 1 : fraction des articles publiés dans les principales revues de physique des particules et qui ont été aussi soumis sur arXiv.org

Même à l'ère des revues électroniques, par conséquent, les prépublications conservent pleinement leur importance, aujourd'hui arXiv contient la vaste majorité des prépublications (dans la plupart des cas incluant les modifications engendrées par le processus de publication) dans le domaine, cela signifie que presque toute la littérature de la discipline est librement accessible sur le Web. Il est essentiel de préciser que cette situation s'est créée sans aucun mandat : les scientifiques ont développé ce schéma parce qu'il répondait à leur propre besoin, comme nous allons le voir dans la prochaine section.

NÉCESSITÉ D'UNE COMMUNICATION RAPIDE ET IMMÉDIATE

Presque toute la littérature produite en physique des particules est disponible sur arXiv. Mais pourquoi est-ce une réalité pour la physique des hautes énergies, alors que dans certaines disciplines, les dépôts disciplinaires ou institutionnels sont à peine remplis ?

Une étude [Gentil-Beccot *et al.*, *in press*] a été réalisée en 2009 sur l'utilisation réelle d'information dans la communauté en utilisant les citations et les *logs* provenant de SPIRES. Nous voyons dans la figure 2 l'une des conclusions de cette étude. Les articles publiés dans cinq revues pendant l'année 2005 ont été divisés en deux échantillons, ceux qui ont été soumis sur arXiv et ceux qui n'ont pas été soumis sur arXiv, nous voyons

clairement que les articles soumis sur arXiv commencent à collecter les citations bien avant leur publication ; à la date de publication, ils ont en effet déjà acquis 20 % du total des citations collectées 24 mois après la publication. Cela s'explique par le fait que les auteurs lisent la prépublication dès qu'elle est disponible sur arXiv, plusieurs mois avant sa publication. Il est à noter que dans la figure 2, les citations issues des prépublications ont été prises en compte au même titre que les citations issues des versions publiées. La rapidité de croissance de la courbe haute s'explique donc, puisqu'il n'y a aucun délai de publication (ni pour le document cité, ou pour celui qui cite). Il y a donc un immense avantage pour les scientifiques à utiliser arXiv : la vitesse de diffusion de l'information. Et, de la même manière que dans les années 1960, les scientifiques envoyaient par la poste leurs prépublications pour diffuser leurs résultats de recherche aussi vite que possible, aujourd'hui, ils utilisent arXiv pour les mêmes raisons, avec, évidemment, une efficacité accrue.

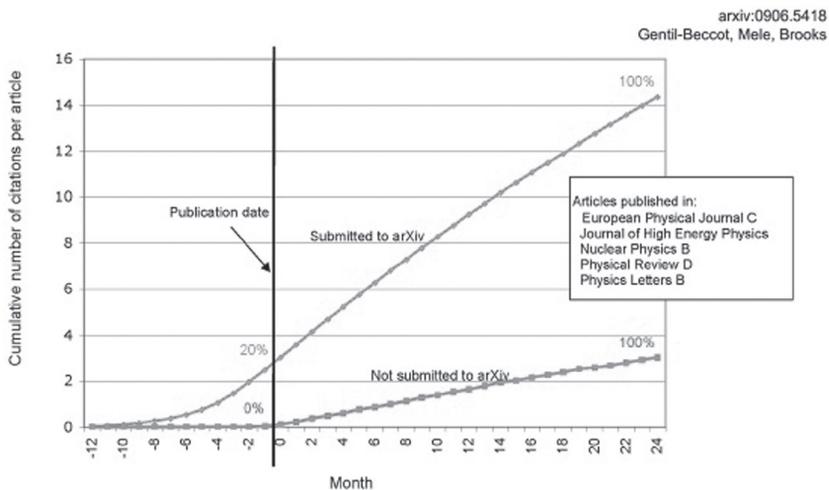


Figure 2 : cumul des citations par article et par mois en fonction de la date de citation par rapport à la date de publication. Les données concernent 4 839 articles de 5 revues majeures en physique des particules : *Journal of High-Energy Physics* (SISSA/Springer), *European Physical Journal C* (Springer), *Physics Letters B* (Elsevier), *Nuclear Physics B* (Elsevier) et *Physical Review D* (APS), au cours de l'année 2005.

L'analyse des citations présentée ci-dessus permet de visualiser la rapidité de la communication scientifique dans la physique des hautes énergies, cependant, elle ne peut être lue que comme une « interprétation » des habitudes de lecture réelles. En revanche, l'analyse des logs de SPIRES donne des informations encore plus concrètes sur les usages de lecture des physiciens des particules.

Une enquête [Gentil-Beccot *et al.*, 2009] a montré qu'environ 50 % des scientifiques utilisent SPIRES pour une recherche bibliographique (figure 3).

Par conséquent, l'analyse des logs issus de SPIRES devrait donner une représentation claire des habitudes de lecture de la communauté. Les logs ont été récoltés pendant un mois, en octobre 2008 (30 000 clics) [Gentil-Beccot *et al.*, *in press*]. L'étude s'est limitée aux clics provenant d'articles offrant à la fois un lien vers le texte intégral sur arXiv et vers la version publiée sur le site de l'éditeur. Nous avons découvert que dans 82 % des cas, arXiv était préféré. L'enquête de 2007 avait également montré que 40 % des scientifiques allaient directement sur arXiv sans passer par SPIRES, ils ne sont donc pas comptés dans les données mentionnées ici. La prédominance d'arXiv sur les versions publiées est donc très vraisemblablement plus élevée que ce qui est montré ici.

Plusieurs caractéristiques typiques de la communauté contribuent à expliquer ce résultat. Une des raisons principales est que, dans la plupart des cas, l'auteur resoumet sur arXiv une version révisée de sa prépublication, incluant les corrections apportées par le processus de révision lors de la publication. Par ailleurs, arXiv offre un accès gratuit à son contenu, alors que la version publiée sur le site de la revue est souvent payante. Les chercheurs privilégient donc un accès direct et libre par définition à un accès qui peut leur être refusé car la bibliothèque n'a pas d'abonnement au journal ou parce qu'ils se trouvent en dehors de l'institut. Ces données confirment un résultat de l'enquête faite en 2007 : les répondants devaient indiquer quel système ils utilisaient le plus pour la recherche de prépublications et pour la recherche d'articles publiés. Les résultats ont montré que les habitudes ne changeaient pas dans un cas ou dans l'autre. Lorsque les chercheurs ont besoin d'accéder à la littérature « classique », ils utilisent donc les mêmes systèmes qui ont été initialement conçus pour recenser la littérature grise dans le domaine.

Les revues scientifiques ont donc perdu leur rôle de fournisseurs d'information et de mode de communication scientifique. Toutefois, ces revues continuent à jouer un rôle indispensable, en fournissant un contrôle

de qualité indépendant qui est essentiel dans ce domaine comme dans de nombreuses autres disciplines scientifiques. En effet, l'évaluation des chercheurs (principalement les jeunes) repose encore principalement sur leurs publications dans les revues prestigieuses à comité de lecture. Le rôle que les revues scientifiques de physique des particules conservent donc aujourd'hui est l'enregistrement et surtout la garantie d'une révision scientifique de qualité. Le projet SCOAP³ [Bianco *et al.* ; SCOAP³, 2010] vise à convertir toutes les revues en physique des particules en libre accès, selon un modèle dans lequel la communauté financerait le peer-review des revues plutôt que l'accès proprement dit aux articles.

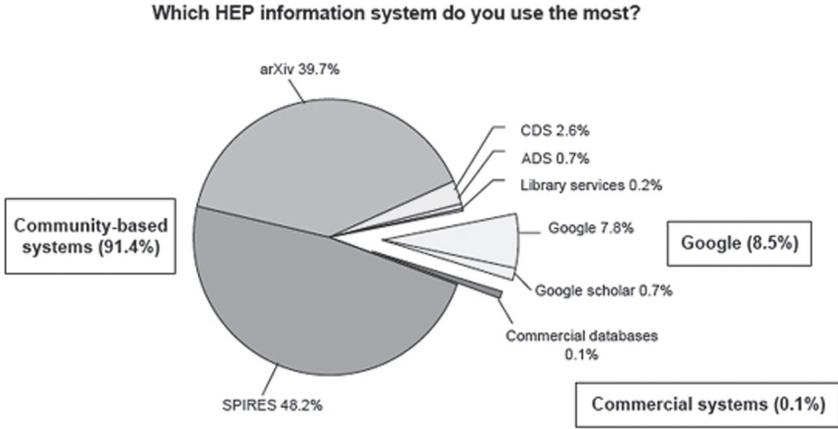
Si le schéma de communication en physique des particules s'est développé ainsi, c'est donc qu'il a suivi les besoins premiers des utilisateurs, mais il a aussi su tirer parti des évolutions technologiques, et plutôt que de résister, il s'est adapté en fonction.

DES OUTILS EN CONSTANTE ÉVOLUTION

 Grâce à l'association d'arXiv et de SPIRES, les scientifiques ont accès à l'ensemble de la littérature en physique des particules, soit par la pré-publication soit par la version publiée, SPIRES fournit des métadonnées détaillées et les références de publication, et arXiv donne accès au texte intégral de presque tous les articles qui sont ensuite publiés dans les revues. Comme nous l'avons mentionné dans la section précédente (figure 3), l'enquête a démontré qu'arXiv et SPIRES sont les deux principaux outils utilisés par la communauté. 91,4% des personnes interrogées préférèrent ces services à tout autre outil. Ce n'est pas un résultat surprenant, car ces systèmes ont été créés par des acteurs de la communauté pour répondre à ses propres besoins, et ces outils se sont adaptés au fil du temps.

D'autre part, 9% des personnes interrogées déclarent utiliser Google ou Google Scholar. L'enquête a également montré que l'utilisation de Google tend à augmenter chez les plus jeunes. Ceci est un reflet de l'augmentation progressive dans la communauté de chercheurs appartenant à une tranche d'âge qui a été formée à la recherche sur Internet avant leurs études universitaires et leur carrière académique, par opposition à des chercheurs qui ont commencé à utiliser des moteurs de recherche au cours de leur activité professionnelle. Il faut ici noter que Google n'est souvent en fait qu'une passerelle vers d'autres sources. En effet, les utilisateurs de Google trouvent le document qu'ils cherchent parce que des ressources comme

arXiv et SPIRES sont indexées par le moteur de recherche. Ce résultat ne peut donc être considéré comme une preuve que Google devient plus important que les outils développés par la communauté, en revanche, il montre clairement le besoin croissant des utilisateurs d'avoir un accès unique à leurs ressources d'information.



arXiv:0804.2701v2

Figure 3 : les ressources d'information préférées des chercheurs en physique des particules

L'enquête a montré par ailleurs que les utilisateurs ont de plus en plus besoin d'avoir un accès plus facile et plus large au contenu, tel que les présentations données aux conférences, la littérature « encore plus grise » (les thèses par exemple) ainsi que les données qui se trouvent derrière les tableaux et les graphiques. L'accès au texte intégral ne suffit plus, les chercheurs ont besoin d'accéder à une information plus complète, plus profonde et plus interconnectée.

Dans cette même étude, on a pu observer que les technologies Web 2.0 prenaient une place non négligeable parmi les exigences des utilisateurs, presque 50 % des personnes qui ont répondu pensent que la « recommandation d'articles », c'est un outil important. Une autre question a tenté d'évaluer le potentiel de la mise en œuvre de fonctionnalités Web 2.0 qui permettraient d'interagir mieux avec l'utilisateur comme le temps que serait prêt à passer un chercheur à commenter/tagger du contenu. 63 % des répondants ont affirmé qu'ils seraient prêts à y passer entre cinq minutes par jour et une heure par semaine, ce qui montre qu'il y a un

immense potentiel d'enrichissement de l'information provenant directement de l'utilisateur. Cette question est essentielle, cette indexation de masse pourrait aider les futurs systèmes d'information à être toujours plus exacts et performants. Nous ne savons pas encore comment cela va évoluer, comment la communauté va utiliser ces nouveaux outils. Mais, ce nouveau type d'information et d'interaction avec l'utilisateur constituera un nouveau défi dans le futur car cela sera sans aucun doute un contenu scientifique qui aura une valeur en tant que telle.

L'enquête réalisée et les discussions entre les quatre plus grands laboratoires de recherche en physique des hautes énergies (CERN, DESY, Fermilab et SLAC), en synergie avec d'autres partenaires (notamment arXiv) et dans un dialogue continu avec les grands éditeurs dans le domaine, ont engendré l'idée d'un nouveau système d'information en HEP. Ce nouveau système est à la croisée des chemins entre l'actuelle base de données SPIRES et une plateforme moderne, Invenio, logiciel open source de bibliothèque numérique [CDS Invenio, 2010]. Ce nouveau système d'information, INSPIRE [INSPIRE, 2010], est développé par une collaboration entre les quatre principaux laboratoires de recherche en physique des particules mentionnés ci-dessus. Il intègre le contenu des bases de données présentes pour accueillir l'ensemble des métadonnées et le texte intégral de toutes les publications en libre accès, passées et futures, comprenant de nouveaux contenus comme les présentations de conférences. Ce nouveau système est appelé à devenir le point d'entrée unique pour l'ensemble de la communauté à tous les documents en physique des particules.

CONCLUSION

+++++

La littérature grise est devenue le canal de communication le plus important parce que les physiciens des particules avaient besoin d'un accès immédiat à l'information. C'est encore le cas, et c'est pourquoi la communauté développe de nouveaux systèmes qui rendent l'accès à l'information encore plus rapide et exhaustif. Mais les besoins évoluent et la communauté doit maintenant aller plus loin : l'accès libre au texte intégral ne suffit plus, les scientifiques veulent une couverture encore plus large et profonde du contenu. Ils ont besoin d'interagir plus directement avec toutes ces ressources. Ce sont tous les défis auxquels le nouveau système d'information HEP INSPIRE doit faire face. Une autre question parallèle est bien sûr l'évolution du rôle de la littérature publiée, qui ne sert plus

directement à la communication scientifique, même si son rôle d'enregistrement et de validation de la qualité ne peut être remis en cause dans le panorama scientifique actuel.

Mais une constante reste dans tous ces mouvements, la communauté elle-même prend la tête de son évolution, guidée par ses propres usages. Par ailleurs, les besoins ne cessent pas d'évoluer avec l'accomplissement des projets, et il sera nécessaire dans le futur de questionner, de mesurer et d'évaluer – encore et toujours – pour adapter au mieux les systèmes qui sont dessinés aujourd'hui.

RÉFÉRENCES BIBLIOGRAPHIQUES

+++++

Addis Louise, *Brief and Biased History of Preprint and Database Activities at the SLAC Library, 1962-1994*, 1994.

[En ligne] < <http://www.slac.stanford.edu/spires/papers/history.html> > (consulté le 24 février 2010).

ArXiv, site Web, [En ligne] < <http://www.arXiv.org> > (consulté le 24 février 2010).

Aymar Robert, "Scholarly communication in high-energy physics: past, present and future innovations", *European Review*, vol. 17, 2009, pp. 33-51, CERN-OPEN-2008-015.

Bianco Stefano *et al*, *Towards Open Access Publishing in High Energy Physics: Report of the SCOAP³ Working Party*, 2007.

[En ligne] < <http://scoap3.org/files/Scoap3WPReport.pdf> > (consulté le 24 février 2010).

CDS Invenio, site Web. [En ligne] < <http://cdsware.cern.ch/invenio/index.html> > (consulté le 24 février 2010).

Gentil-Beccot Anne, Mele Salvatore, Holtkamp Annette, O'Connell Heath et Brooks Travis C., "Information resources in high-energy physics: surveying the present landscape and charting the future course", *Journal of the American Society for Information Science and Technology*, vol. 60, 2009, pp. 150-160, arXiv : 0804.2701v2.

Gentil-Beccot Anne, Mele Salvatore et Brooks Travis C., "Citing and reading behaviours in high-energy physics. How a community stopped worrying

about journals and learned to love repositories”, *Scientometrics*, 84, 2010, pp. 345-355, arXiv : 0906.5418.

Ginsparg Paul, “First steps towards electronic research communication”, *Computers in Physics*, vol. 8, 1994, pp. 390-396.

Goldschmidt-Clermont Luisa, “Communication patterns in high-energy-physics”, *High-Energy Physics Libraries Webzine*, n° 6, 2002.

[En ligne] < <http://library.web.cern.ch/library/Webzine/6/papers/1/> > (consulté le 24 février 2010).

Heuer Rolf-Dieter, Holtkamp Annette et Mele Salvatore, “Innovation in scholarly communication: vision and projects from high-energy physics”. *Information Services & Use*, vol. 28, 2003, pp. 83-96, arXiv : 0805.2739.

INSPIRE. [En ligne] < <http://inspirehep.net/> > (consulté le 24 février 2010).

Kreitz Patricia A. et Brooks Travis C., “Subject access through community partnerships: a case study”, *Science & Technology Libraries*, vol. 24, 2003, pp. 153-172, arXiv : physics/0309027.

Kunz Paul *et al*, *The Early World Wide Web at SLAC*.

[En ligne] < <http://www.slac.stanford.edu/history/earlyweb/history.shtml> > (consulté le 24 février 2010).

Mele Salvatore, Dallman David, Vigen Jens et Yeomans Joanne, “Quantitative analysis of the publishing landscape in high-energy physics”, *Journal of High-Energy Physics*, vol. 12, 2006, S01, arXiv : cs.DL/0611130.

SCOAP³, site Web. [En ligne] < <http://www.scoap3.org/> > (consulté le 24 février 2010).

SPIRES, site Web. [En ligne] < <http://www.slac.stanford.edu/spires> > (consulté le 24 février 2010).