

La Théorie sur la "voie verte" de l'Open Access

Jocelyne Jerdelet,
jocelyne.jerdelet@cern.ch
CERN, Genève

Sandrine Reyes,
sandrine.reyes@cern.ch
CERN, Genève

Résumé

Le domaine de la physique des particules a toujours œuvré pour la mise à disposition des résultats de recherche par la diffusion des pré-tirages. Ceux-ci étaient déjà distribués sous forme de document papier bien avant l'ère d'Internet. Depuis la Convention du CERN en 1953, les pré-tirages ou prépublications rédigés par les physiciens théoriciens ont été conservés, classés puis catalogués dans la base de données du Service d'Information Scientifique (SIS) et, par la suite, numérisés. Étudiée de façon rétrospective, cette littérature grise aujourd'hui stockée sur le serveur du CERN, est mise en ligne et devient ainsi librement accessible à toute la communauté scientifique. On parle de « voie verte » de l'Open Access.

Mots-clés

Littérature grise, Pré-tirage, Prépublication, Libre accès, Dépôt institutionnel, Serveur, Voie Verte

1. Le libre accès – Open Access (OA)

Le mouvement du libre accès désigne l'ensemble des initiatives prises pour une mise à disposition des résultats de la recherche au plus grand nombre, sans restriction d'accès (libre et gratuit) pour les lecteurs, que ce soit par l'auto-archivage ou par la création de revues en libre accès.

- L'auto-archivage sur un serveur institutionnel ou sur une page personnelle, communément appelé "la voie verte", consiste à déposer par les auteurs une copie de leur article publié dans des archives électroniques en libre accès, avec l'accord de l'éditeur.
- La publication dans des revues scientifiques en libre accès communément appelée "la voie or", doit satisfaire à des exigences qualité, garanties par un comité de lecture. Son mode de financement permet une diffusion sans restriction d'accès et d'utilisation.

Le développement d'Internet, l'arrivée du World Wide Web (WWW), l'explosion des documents électroniques, l'augmentation du prix des abonnements des périodiques et le besoin d'accéder plus facilement et plus rapidement à l'information, sont les principaux facteurs qui ont contribué au développement de l'Open Access (OA).

Ce mouvement prend son essor en décembre 2001 à Budapest, lors d'une rencontre restreinte parrainée par l'Institut pour la Société Ouverte (Open Society Institute, OSI) [1], qui vise à harmoniser les différentes initiatives jusque là disparates.

Par la suite d'autres rencontres ont eu lieu, donnant naissance à trois textes fondateurs appelés "3B".

- **« Budapest Open Access Initiative (BOAI) » [2] - Février 2002** L'initiative de Budapest recommande deux stratégies pour le libre accès : La « voie verte » et la « voie or ».
- **« Bethesda Statement on Open Access Publishing (OAP) » [3] – avril 2003** La déclaration de Bethesda définit provisoirement la publication en libre accès.
- **« Berlin Declaration Open Access to knowledge in the Sciences and Humanities » [4] – octobre 2003** La déclaration de Berlin sur le libre accès à la connaissance en sciences exactes, en sciences de la vie, sciences humaines et sociales s'engage à promouvoir et à soutenir la diffusion gratuite de la connaissance via Internet. Cette déclaration est alors signée par le CERN.

2. Open Access au CERN ⁽¹⁾

Depuis sa création, le CERN a toujours activement soutenu les principes de l'Open Access. Sa Convention, adoptée le 1er juillet 1953 par 12 Etats Membres ⁽²⁾, stipule que « les résultats de ses travaux expérimentaux et théoriques [de l'Organisation] sont publiés ou de toute autre façon rendus généralement accessibles » [5].

En 1989, Tim-Berners Lee, scientifique au CERN, invente le World Wide Web [6] qui popularise l'Internet et ouvre la voie verte de l'Open Access.

En 1991, Paul Ginsparg, physicien à Los Alamos, met en place le premier serveur de documents électroniques, nommé hep-th (High-Energy Physics – Theory), première archive

en libre accès connue aujourd'hui sous le nom d'arXiv⁽³⁾ ou encore le serveur de Cornell car maintenu par l'Université de Cornell.

En 1993, suivant le même exemple, le CERN ouvre son propre serveur de documents électroniques en libre accès (CDS⁽⁴⁾ - dépôt institutionnel) adapté aux besoins des chercheurs et des documentalistes. Les notices bibliographiques sont consultables via Internet sur ce catalogue en ligne, ainsi que le document accessible grâce à un lien hypertexte.

En 2005, le Comité de Direction du CERN adopte une nouvelle politique de libre accès concernant tous les résultats du Laboratoire, définie dans le document « Continuing CERN action on Open Access » [7]. Les auteurs sont encouragés à publier leurs travaux dans des revues scientifiques en libre accès.

En novembre 2006, les représentants des principales agences de financement européennes pour la Physique des Particules, de consortiums de bibliothèques et de la communauté scientifique se réunissent au CERN pour le lancement du projet SCOAP3⁽⁵⁾ (Sponsoring Consortium for Open Access Publishing in Particle Physics). Son but est de changer le mode de financement des futures publications en Physique des Particules. L'accès libre à ces futurs articles serait financé, non plus par les abonnements aux journaux souscrits annuellement par chaque bibliothèque mais par un fond commun constitué grâce au transfert de ces fonds. Déjà soutenu par plusieurs bibliothèques et agences de financement européennes, SCOAP3 s'organise maintenant à un niveau mondial.

3. Le Service d'Information Scientifique (SIS) du CERN et l'Open Access

Le Service d'Information Scientifique du CERN, créé en 1955, a pour mission principale d'acquérir et de gérer l'information concernant les travaux du CERN et de la diffuser à la communauté scientifique, selon la « Circulaire Opérationnelle N°6 » [8].

En 1982, le SIS informatise son catalogue par la mise en place d'un système de gestion des données bibliographiques appelé ISIS⁽⁶⁾ fourni par l'UNESCO⁽⁷⁾.

En 1989, ISIS est remplacé par le logiciel ALEPH⁽⁸⁾ [ExLibris, Israel]. Ce système intégré, plus adapté aux besoins de la bibliothèque du CERN, propose différents modules utiles pour traiter tous types de documents (Catalogage, Gestion des copies, Prêt, Circulation, Acquisition, ...).

L'arrivée du Web, la démocratisation d'Internet et les avancées technologiques telles que la numérisation de documents et la mise en place de serveurs de documents électroniques en libre accès ont bouleversé les milieux scientifiques et bibliothéconomiques.

Le SIS reçoit régulièrement, via des listes de diffusion, des documents papier, rédigés par des scientifiques du CERN et d'autres Instituts de recherches. Ces pré-tirages sont catalogués puis numérisés. Plus tard, la soumission électronique par l'auteur se substitue au catalogage de la bibliothèque. Très vite, le support papier disparaît au profit d'un document électronique. C'est alors que débute l'ère de la bibliothèque numérique.

Le SIS s'adapte et développe un programme d'importation de notices bibliographiques provenant de bases de données des Instituts collaborant avec le CERN (SLAC⁽⁹⁾, DESY⁽¹⁰⁾...), de serveurs de pré-tirages (arXiv) et de pages Web (autres instituts) pour remplacer les listes de diffusions et pour offrir aux chercheurs un accès quasi unique à

l'information. Ainsi les résultats de recherche du CERN et d'autres instituts sont mis à la disposition de la communauté scientifique sur l'interface CDS à travers de nombreuses notices bibliographiques accompagnées de leurs fichiers électroniques.

En 2006, le Service d'Information Scientifique s'investit d'avantage dans la "voie verte de l'Open Access" en mettant en ligne la série complète des documents consacrés à la physique théorique (TH) du Laboratoire.

4. Choix de la Série Théorie

Le choix s'est porté sur les pré-tirages de la division Théorie car cette collection est la plus ancienne (juillet 1953 à décembre 2006) et la plus prolifique du CERN, avec un total d'environ 12000 documents.

Début 2006, à la demande de la Section Gestion des Documents du SIS, ces pré-tirages sont versés par le Secrétariat de la Théorie aux Archives Historiques et Scientifiques.

Depuis 1958, cette documentation fut répertoriée manuellement par liste dans des registres tenus par le Secrétariat TH et identifiée par un numéro de rapport unique.

5. Historique du catalogage de la série TH

Ces pré-tirages sous forme papier ne furent catalogués qu'à partir de 1982. Les données furent ensuite converties en format ISBD(M) ⁽¹¹⁾ dans le système de gestion informatique de données ALEPH. La numérisation, quant à elle, débute en 1990 pour les nouveaux documents.

Dès 1991, les auteurs de la théorie prennent l'habitude de soumettre leurs documents sur le serveur arXiv. A partir de 1994, pour rendre disponible cette documentation à travers le serveur du CERN, le SIS importe les métadonnées d'arXiv. Simultanément, l'institut KEK ⁽¹²⁾ au Japon réalise un travail considérable de numérisation de pré-tirages en physique dont certains du CERN. Le SIS obtient alors l'autorisation de KEK de lier leurs numérisations aux notices bibliographiques du Serveur de Documents du CERN (CDS).

6. L'objectif du projet « voie verte de l'Open Access »

Après l'analyse historique du catalogage de la collection TH, les objectifs du projet sont définis selon les critères de la voie verte de l'Open Access, et aussi selon les besoins des services du SIS et des Archives Historiques et Scientifiques :

- Pré-tirage catalogué
- Recherche de la référence de publication et mise à jour de la notice bibliographique
- Texte intégral en accès libre sur CDS
- Copie papier conservée aux Archives

7. Tâches et difficultés rencontrées

- **Catalogage** : après une première analyse de la collection, 9457 notices sont déjà répertoriées dans la base de données du SIS mais beaucoup restent à traiter, notamment les 2000 premiers documents. La perte de certaines données (telles que les numéros de rapports) suite aux diverses conversions informatiques, implique, avant tout catalogage de nouvelle notice, une recherche bibliographique par titre pour tous

les numéros de rapports manquants. De plus avec l'évolution des normes bibliographiques (depuis ISBD(M) jusqu'au format MARC21 ⁽¹³⁾), l'homogénéité et la qualité du catalogage restent à vérifier sur l'ensemble des notices. Le système manuel de numérotation jusqu'en 2004 a rendu le travail du SIS plus fastidieux, car dépendant d'une recherche également manuelle.

- **Numérisation** : après la finalisation du catalogage des nouvelles notices, une analyse met en évidence qu'environ 50% de la collection possède un lien vers un texte intégral provenant du CERN, du serveur d'ArXiv, de KEK ou autres... Avant toute numérisation de document, une recherche dans la base de données de KEK permet une nouvelle fois l'importation de documents numérisés. Cette tâche achevée, 2527 documents restants doivent être équipés d'un code-barres, code unique d'identification sur le serveur du CERN, en vue d'être scannés.
- **Standardisation et extraction de données** : pour plus d'efficacité, les liens vers les fichiers des documents à numériser par le SIS sont ajoutés globalement dans les notices grâce à une extraction de données.
- **Vérification des liens URL** : une fois la numérisation terminée, un contrôle final de tous les liens URL est effectué sur l'ensemble de la collection pour détecter d'éventuels fichiers corrompus.
- **Recherche de Références de Publication** : afin d'enrichir le catalogue, des recherches sont entreprises dans différentes sources, bases de données ou moteurs de recherche tels que INSPEC ⁽¹⁴⁾, SPIRES ⁽¹⁵⁾, GOOGLE ⁽¹⁶⁾ pour enrichir au maximum les notices bibliographiques avec les références de publication (journaux, comptes-rendus de conférences).
- **Enquête auprès des auteurs** : à la fin de l'exercice, 267 références de pré-tirages ne sont pas disponibles en Open Access car ces documents n'ont jamais été versés au Secrétariat TH. Une enquête auprès des auteurs est menée en vue d'obtenir un exemplaire papier, un fichier électronique ou une référence de publication. Sur 167 courriers électroniques envoyés, 29 auteurs ont répondu positivement.

La difficulté résulte essentiellement au fait que beaucoup de ces pré-tirages datent des premières années du CERN et qu'il n'est pas aisé de rentrer en contact avec ces auteurs.

8. Evaluation finale de l'analyse

Depuis le début du CERN, la production de pré-tirages de la division Théorie a été en constante augmentation. Aujourd'hui, cette collection représente 11780 notices disponibles sur CDS, parmi lesquelles :

- 9457 Notices initialement existantes et vérifiées
- 928 Nouvelles notices cataloguées
- 1395 Notices recherchées et identifiées comme série TH

Une grande majorité de ces documents, soit 10145, sont des articles publiés dans des journaux scientifiques ou dans des comptes-rendus de conférences, 1579 n'ont jamais été publiés et le reste étant des documents tels que des rapports, des livres, des comptes-rendus de conférences et des thèses.

Actuellement, suite aux différentes recherches, enquêtes et numérisations effectuées, 98% de la collection Théorie est disponible sur CDS en Open Access :

<http://cdsweb.cern.ch/collection/CERN-TH>

L'analyse complète de la collection Théorie a nécessité 3520 heures de temps d'étude, soit 2 Equivalents Plein Temps (EPT) et 250 heures de numérisation (2527 documents), soit 0.15 Equivalent Plein Temps (EPT).

9. Conclusion

La diffusion des résultats de recherche scientifique, d'abord effectuée par des Sociétés Savantes, est aujourd'hui diffusée, généralement, par des journaux à la publication parfois tardive et aux abonnements onéreux. Elle reste donc réservée à un ensemble de chercheurs privilégiés. Le choix du SIS, de favoriser l'accès libre aux prépublications scientifiques, résulte d'une volonté de transmettre par un moyen efficace et économique la connaissance et l'évolution des progrès de la science. Cette mise à disposition gratuite et plusieurs mois avant la publication de ces pré-tirages, représente un intérêt particulier, notamment en Physique des Particules.

Commencé en 2006, ce projet de mise en ligne dans CDS et de conservation des documents aux Archives Historiques et Scientifiques du CERN doit s'appliquer à la totalité des collections de pré-tirages du CERN (rétrospectives et courantes). Ainsi ces prépublications pourront être réutilisées, copiées, téléchargées, imprimées, archivées, consultées à distance, offrant aux utilisateurs rapidité, fiabilité, facilité d'utilisation et gratuité de l'information.

Le développement et la généralisation de la voie verte de l'Open Access au sein des dépôts institutionnels permettraient d'assurer, préserver et pérenniser l'accès à la totalité des résultats scientifiques, sans limitation d'utilisation, sous l'unique contrôle des instituts et ceci dans un esprit d'économie non négligeable.

NOTES

- (1) CERN : Organisation Européenne pour la Recherche Nucléaire - <http://cern.ch>
- (2) 12 Etats Membres en 1953 : Belgique, Danemark, France, Grèce, Italie, Norvège, Pays-Bas, République Fédérale Allemande, Royaume-Uni de Grande Bretagne et d'Irlande du Nord, Suède, Suisse, Yougoslavie.
- (3) Serveur arXiv - <http://arxiv.org/>
- (4) CDS : CERN Document Server - <http://cdsweb.cern.ch/>
- (5) SCOAP3 - <http://www.scoap3.org/>
- (6) ISIS : "Integrated Set for Information" est un système de gestion de catalogage.
- (7) UNESCO : United Nations Educational, Scientific and Cultural Organization - Organisation des Nations Unies pour l'Education, la Science et la Culture.
- (8) ALEPH : "Automated Library Expandable Program" est un logiciel de gestion intégré des bibliothèques, développé par l'université de Jérusalem et produit par la société Ex-libris.
- (9) SLAC : Stanford Linear Accelerator Center, Université de Stanford, Californie, USA.
- (10) DESY : Deutsches Elektronen-Synchrotron, Hamburg, Allemagne.
- (11) ISBD(M) : International Standard Bibliographic Description (description bibliographique internationale normalisée). (Il s'agit d'un ensemble de normes internationales de description de catalogage définies par l'IFLA, Fédération Internationale des Associations de Bibliothécaires et d'Institutions).
- (12) KEK : High Energy Accelerator Research Organization, Tsukuba, Japan.
- (13) MARC21 : « Machine – Readable Cataloguing » Format standard pour le stockage et l'échange de notices bibliographiques maintenu par « The Library of Congress » - Washington DC.
- (14) INSPEC - <http://www.theiet.org/publishing/inspec/>
- (15) SPIRES - <http://www.slac.stanford.edu/spires/>
- (16) GOOGLE - <http://www.google.ch/>

BIBLIOGRAPHIE

- [1] Open Society Institute (OSI) & Soros Foundation Network - <http://www.soros.org/about>
- [2] Budapest Open Access Initiative (BOAI), février 2002 - <http://www.soros.org/openaccess/fr/read.shtml>
- [3] Bethesda Statement on Open Access Publishing (OAP), avril 2003 - <http://www.earlham.edu/~peters/fos/bethesda.htm>
- [4] Berlin Declaration Open Access to knowledge in the Sciences and Humanities - <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

- [5] La Convention pour l'établissement d'une Organisation Européenne pour la Recherche Nucléaire (CERN), adoptée le 1er juillet 1953 par 12 Etats Membres - <http://cdsweb.cern.ch/record/330625>
- [6] Histoire du World Wide Web - <http://info.cern.ch/default-fr.html>
- [7] « Continuing CERN action on Open Access » CERN-OPEN-2005-006 – Altarelli, Guido - <http://cdsweb.cern.ch/record/828991>
- [8] Circulaire Opérationnelle N°6 - <http://open-access.web.cern.ch/Open-Access/oc-06.pdf>