Rapport de stage septembre 2002

DESS Ingénierie documentaire

Evolution des serveurs de documentation scientifique et technique en ligne dans la perspective de l'auto archivage

Le cas du CERN

Analyse du contexte et propositions appliquées.

Sarah Faraud

Sous la direction de Corrado Pettenati
Chef de la bibliothèque du CERN

enssib
école nationale supérieure des sciences de l'information et des bibliothèques

Evolution des serveurs de documentation scientifique et technique en ligne dans la perspective de l'auto archivage

Le cas du CERN

Descripteurs : Auto archivage, documentation scientifique, physique des hautes énergies, OAI, pré publications, technologies de l'information.

Résumé : Le service de documentation scientifique du CERN est au cœur des évolutions récentes de la diffusion scientifique. Le mouvement de l'OAI et le développement de l'auto archivage sont des perspectives dans lesquelles le CERN souhaite s'engager. Une analyse de ces bouleversements et de leur validité ainsi que des propositions concrètes étaient nécessaires avant d'adopter toute stratégie nouvelle.

Evolution of online scientific information servers through the prism of self-archiving. The CERN : a case study.

Keywords : Self-archiving, scientific documents, preprints,, high energy physics, OAI, information technology

Abstract : The CERN information service has always been involved with the latest developments in information technologies. OAI and self-archiving are notions that are now prevailing and the launching the new CDS server needed to be done with a precise analysis of the stakes and issues in these fields. In the same perspective, a set of new features has been submitted to the service for approval.

Remerciements

A Corrado Pettenati pour m'avoir épaulée et fait confiance pendant ce stage.

A Jens Vigen pour sa disponibilité , son entrain et son indéfectible bonne humeur

A Tulio Basaglia pour son sourire et son humour

A Jean-Philippe Schmidt pour se conseils avisés

Aux membres de la bibliothèque du CERN pour m'avoir acceptée et intégrée à la vie du service.

A Peter Suber pour son aide.

A mes parents et à mes amis pour tout le reste

A mon cousin et à sa famille pour leur accueil

Sommaire

Rapport de stage septembre 2002 I

SOMMAIRE..... 4

INTRODUCTION 6

LE CERN : CONSEIL EUROPEEN POUR LA RECHERCHE NUCLEAIRE . 8

1. Historique 8
   - 1.1. La gestation du projet 8
   - 1.2. Du premier accélérateur aux premières expériences 9
   - 1.3. Les évolutions récentes 11

2. La physique nucléaire et ses instruments 12
   - 2.1. Les accélérateurs 12
   - 2.2. Le LEP 13

3. Les divisions du CERN 15

4. La bibliothèque du CERN 16
   - 4.1. Organisation du service 16
     - 4.1.1. Mission 16
     - 4.1.2. Le personnel 16
   - 4.2. Le Cern Document Server, service en ligne de la bibliothèque du CERN 17
     - 4.2.1. Les collections 17
     - 4.2.2. Le moteur de recherche 18
     - 4.2.3. Les services de recherche 19
     - 4.2.4. Les services de soumission 20

MISSION 21

1. Le contexte 21
   - 1.1. L'auto archivage 21
   - 1.2. Le CERN et l'auto archivage 22

2. Méthode de travail 23
   - 2.1. Phase de familiarisation 23
     - 2.1.1. Familiarisation avec le domaine 23
     - 2.1.2. Avec le CERN 23
     - 2.1.3. Avec la thématique du stage 23
   - 2.2. Organisation de la mission et plan de travail 24
     - 2.2.1. Des réunions hebdomadaires avec les tuteurs 24
     - 2.2.2. Des présentations régulières au service 25
   - 2.3. Analyse de l'existant sur CDS 25
     - 2.3.1. L'analyses des tables 25
     - 2.3.2. L'analyse du moteur de recherche 28
   - 2.4. Veille documentaire 30
     - 2.4.1. Les listes et newsletters 30
     - 2.4.2. Les principales sources publiées et les sites 31

2.5. Rencontres, enquête et interview 32
2.5.1. Informaticiens 32
2.5.2. Professionnels de l'information 34

PROPOSITIONS D'AMELIORATION 36

1. CDS et l'auto archivage. 36
1.1. Self-archiving, super archive, institutional repository, trusted digital repository : des notions à élucider. 36
1.1.1. Self archiving, auto-archivage 36
1.1.2. Superarchive et institutional repository 37
1.1.3. Trusted digital repository 39
1.2. CDS : vers des activités de plus en plus diversifiées. 41
1.2.1. Le CERN comme Trusted digital repository 41
1.2.2. CDS une interface pour un service riche 41

2. Valorisation des documents en ligne 42
2.1. De la vérification à la certification 43
2.2. Modifications de l'interface ajout de fonctionnalités 49
2.2.1. Nombre de déchargements 49
2.2.2. Nombre de citations 51
2.2.3. Ajout de fonctionnalités dans le cadre de super archive 52

CONCLUSION 55

BIBLIOGRAPHIE..... 56

ANNEXES 69

Introduction

La communication scientifique se fait principalement par des articles, des conférences, et des rapports internes. Les évolutions récentes des technologies de transmission de l'information sont synonymes de bouleversement dans tous les domaines et touchent tout particulièrement le monde de la recherche scientifique.

Jusqu'alors, l'accès aux articles se faisait par le biais d'abonnements souscrits par les bibliothèques a des prix souvent ahurissants, excédant dans bien des cas les budgets des bibliothèques et en forçant certaines à renoncer à certaines revues. La mise en ligne de journaux gratuits au début des années 1990 , comme *The Public Access Computer Systems Review* ou *Postmodern Culture*, était une réaction contre cet état de fait. Les principales revues scientifiques ne tardèrent pas à offrir des accès en ligne mais les prix ,loin d'être revus à la baisse, continuèrent à croître. De fait, dès le début des années 1990, une alternative gratuite et simple s'est développée dans le domaine de la physique : le serveur ArXiv de Paul Ginsparg dont le but était d'offrir un espace aux chercheurs où déposer leurs articles avant la publication. En 1994, Steven Harnad proposait la notion d'auto archivage (self-archiving) et lançait quelques années plus tard, en 1997, le serveur Cogprint dédié à la psychologie et aux neurosciences. L'idée était dans la droite ligne de ce que faisait Ginsparg, c'est à dire de permettre aux chercheurs de déposer leurs articles avant publication et après publication, et de négocier le droit de mettre la version révisée dans l'archive. Cette idée s'est développée jusqu'à aujourd'hui où elle remporte un vif succès dans le monde entier, dans le cadre technique instauré par l'Initiative pour les Archives Ouvertes (OAI) qui a établit un protocole d'extraction des méta données des articles, utilisable et applicable à toutes les archives de ce type, les rendant par la même inter opérables.

Ces alternatives ont entraîné une simplification de l'accès à un nombre de documents en augmentation constante. Cependant, le grand nombre de documents ne va pas sans poser problème à plusieurs niveaux : leur gestion par des outils de recherche performants est complexe, leur conservation à long terme doit être assurée quel que soit le format d'origine, une signalétique permettant d'identifier les différents types de documents doit être mise en place. Toutes ces évolutions doivent s'intégrer dans les nouvelles

perspectives de l'auto archivage, amenant inévitablement le service de documentation scientifique du CERN à se transformer afin de répondre aux mieux aux attentes de la communauté scientifique. C'est dans ce contexte que s'est déroulée ma mission au sein du service de documentation scientifique du CERN entre juin 2002 et fin septembre 2002.

Le CERN : Conseil européen pour la recherche nucléaire

1. Historique

1.1. La gestation du projet

A l'issue de la seconde guerre mondiale, un nouvel esprit de coopération internationale apparut, avec la création des Nations Unies et de ses agences spécialisées. En Europe, l'Organisation pour la Coopération Economique Européenne fut le précurseur de ce qui est aujourd'hui la Communauté Européenne. En plus des conséquences de la guerre proprement dite, le prestige scientifique de l'Europe a beaucoup souffert de l'émigration de scientifiques de premier plan - la "fuite des cerveaux" avait commencé. Pour rétablir l'équilibre et redonner du prestige à la science européenne, le physicien français, lauréat du Prix Nobel, Louis de Broglie propose la création d'un laboratoire scientifique européen à l'occasion de la Conférence Européenne de la Culture à Lausanne en 1949.
A la 5ème conférence générale de l'UNESCO à Florence en 1950, le physicien américain, lauréat du Prix Nobel, Isidore Rabi propose une résolution, unanimement adoptée, autorisant l'UNESCO, "à fournir assistance et à encourager la formation et l'organisation de centres et de laboratoires régionaux afin d'accroître et de rendre la collaboration internationale entre scientifiques plus fructueuse ...". Suite à deux conférences de l'UNESCO en 1952, onze gouvernements européens décident la mise en place, à titre provisoire, d'un "Conseil Européen pour la Recherche Nucléaire" (CERN). Lors d'une réunion à Amsterdam du tout nouveau Conseil du CERN, un site proche de Genève est sélectionné pour accueillir le futur laboratoire. Après ratification initiale de la Convention par ses Etats Membres, l'Organisation Européenne pour la Recherche

Nucléaire voit le jour le 29 septembre 1954. Bien que le CERN "provisoire" soit dissout, l'acronyme, devenu familier, est conservé. Les douze Etats Membres fondateurs sont: la République Fédérale d'Allemagne, la Belgique, le Danemark, la France, la Grèce, l'Italie, la Norvège, les Pays-Bas, le Royaume-Uni, la Suède, la Suisse, et la Yougoslavie. Cette dernière quitte le CERN en 1961. L'Autriche et l'Espagne y adhèrent respectivement en 1959 et 1961. L'Espagne quitte à son tour le CERN en 1969 pour y revenir en 1983. Le Portugal se joint au CERN en 1985, la Finlande et la Pologne en 1991, la Hongrie en 1992, les Républiques Tchèque et Slovaque en 1993 et la Bulgarie en 1999, portant le nombre d'Etats Membres à 20.

1.2. Du premier accélérateur aux premières expériences

En 1957, le premier accélérateur du CERN, un **Synchro-Cyclotron (SC)** à protons de 600 MeV, est mis en service. L'une des premières réussites expérimentales est l'observation, attendue depuis si longtemps, de la désintégration d'un pion en un électron et un neutrino. En 1959, la première grande machine du CERN, le **Synchrotron à Protons (PS)** de 28 GeV, est mise en service et les premières images d'interactions neutrino dans les chambres à bulles du CERN sont observées. La physique des neutrinos tirera un grand bénéfice de l'éjection rapide des protons du synchrotron.

En 1965 un accord, passé avec les autorités françaises, étend le domaine du CERN en territoire français. Le Conseil du CERN approuve la construction des **Anneaux de Stockage à Intersections (ISR)**, premier collisionneur de protons au monde, qui sera construit sur cette extension et qui entrera en service en 1971. C'est en 1967 que le CERN inaugure le **séparateur d'isotopes en ligne ISOLDE** pour l'étude des noyaux à durée de vie très courte. Cette installation de classe mondiale élargit grandement le champ de recherche du CERN. Grâce à un accord spécial entre le CERN et la France, le travail sur la **chambre à bulles à liquides lourds**, Gargamelle, commence. Un accord est passé entre le CERN, la France et l'Allemagne concernant la construction d'une chambre à bulles à hydrogène de 3,7 mètres équipée du plus grand aimant supraconducteur du monde.

En 1971, on donne l'approbation à la construction d'un second laboratoire, adjacent au site existant et qui abritera un **Super Synchrotron à Protons (SPS)** de 7 km de circonférence, prévu pour être initialement exploité à une énergie de 300 GeV. Bien qu'administrativement séparés au début, les deux laboratoires du CERN seront réunis en 1976. Un booster à quatre anneaux de 800 MeV est achevé pour augmenter l'énergie d'injection du PS. Grâce au booster et à un nouveau linac, mis en service en 1978, le PS ne cesse d'accroître son intensité, parvenant même à atteindre mille fois la valeur nominale. Le complexe, unique au monde, d'accélérateurs interconnectés du CERN, avec le PS en son cœur, fournit une variété sans pareille de faisceaux de particules et de possibilités de recherche

Les premières découvertes importantes dues aux expériences réalisées aux ISR montrent que la taille du proton croît à mesure que l'énergie augmente, et les particules émergeant sous des angles prononcés révèlent la diffusion des constituants, profondément enfouis dans les protons. La chambre à bulles Gargamelle, installée dans un faisceau de neutrinos au PS permet au CERN de faire l'une de ses plus importantes découvertes en physique: les neutrinos peuvent interagir avec une autre particule tout en restant des neutrinos. Cette "interaction à courant neutre" constitue une nouvelle avancée, apportant du crédit à la théorie qui essaie d'unifier notre compréhension de la force faible - à la base de phénomènes tels que la radioactivité - et de la force électromagnétique qui nous est familière. Les performances de l'accélérateur ne cessent de s'accroître et l'intensité nominale est rapidement dépassée. Fin 1978, l'énergie de pointe est portée à 500 GeV. Des expériences faites au CERN montrent comment la qualité et l'intensité des faisceaux peuvent être améliorées grâce à la "technique du refroidissement stochastique" proposée au CERN par Simon van der Meer en 1968. La possibilité d'accélérer et de stocker d'intenses faisceaux de particules permet alors de faire le projet hardi de convertir le SPS en un collisionneur protons-antiprotons, en utilisant un **anneau d'Accumulation d'Antiprotons (AA)** basé sur le refroidissement stochastique des antiprotons.

1.3. Les évolutions récentes

En 1989, le **LEP,** Grand Collisionneur Electrons-Positons (LEP) de 27 km de circonférence, le plus grand instrument scientifique jamais construit, avec une énergie initiale de fonctionnement de 50 GeV par faisceau, est mis en service au mois d'août. En octobre, deux mois seulement après les premières collisions dans le LEP, des mesures d'une extrême finesse de la particule Z sont réalisées et montrent que les briques fondamentales, à partir desquelles toute la matière est faite, sont constituées de trois, et seulement trois, familles de particules. Le 13 novembre, le LEP est inauguré officiellement en présence de chefs d'Etat et de ministres de la recherche scientifique.

En 1990, Tim Berners-Lee, en collaboration avec Robert Cailliau, également du CERN, propose un système d'information distribué basé sur l'hypertexte, un moyen d'établir des liens entre des documents stockés dans différents ordinateurs. En cachant la complexité de l'adressage des documents au sein des ordinateurs et des ordinateurs au sein du réseau, derrière des groupes de mots en sur écriture identifiant les documents, il est possible de relier ces informations entre ordinateurs. Le nom "**World-Wide Web**" est adopté, c'est une découverte qui révolutionnera le monde de l'information.

En décembre, les délégués au Conseil du CERN, reconnaissent à l'unanimité que le **Grand Collisionneur de Hadrons (LHC**), qui serait construit dans le tunnel du LEP, serait la meilleure machine pour le futur

Depuis 1989, chaque année apporte sa moisson de succès des expériences LEP. Le résultat le plus remarquable est la précision avec laquelle les paramètres de résonance de la particule Z ont été mesurés: de 1989 à 1993, les **quatre détecteurs LEP - ALEPH, DELPHI, L3 et OPAL** - ont reconstruit plus de 10 millions de désintégrations Z. C'est aussi l'année du 40^{e} anniversaire du CERN. Le Conseil approuve la construction du LHC.

En septembre 1995, une équipe internationale dirigée par Walter Oelert parvient à synthétiser des atomes d'antimatière à partir de leurs antiparticules constituantes. La création d'atomes d'antimatière au CERN ouvre ainsi la porte à l'exploration systématique de l'antimonde. Après avoir apporté des contributions financières

significatives au projet LHC, le Japon devient Etat Observateur auprès du CERN. L'énergie du LEP est augmentée pour permettre la production de paires de particules W. Après avoir accepté de contribuer financièrement de façon significative au LHC, les Etats-Unis deviennent Etat Observateur auprès du CERN.

2. La physique nucléaire et ses instruments

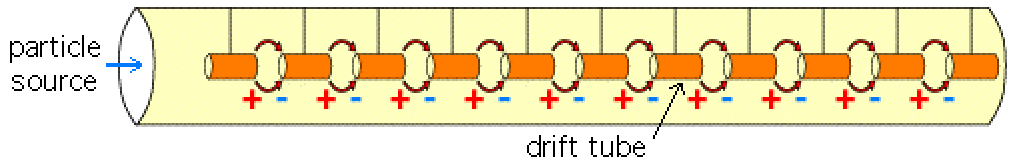
2.1. Les accélérateurs

Le CERN étudie ce qu'est la matière, ce dont elle est faite et les forces qui la maintiennent agglomérée. A l'origine des faisceaux de particules, il y a toujours une **source de particules**. La source de particules la plus simple n'est rien d'autre qu'un filament chauffé, comme les filaments à l'intérieur des ampoules électriques. C'est une source de ce type qui est utilisée dans les téléviseurs. Des électrons chargés négativement s'échappent du filament et accélèrent en direction d'une électrode chargée positivement, qu'ils traversent. Ensuite, sous l'action de champs électromagnétiques, le faisceau balaye l'écran, où il rend lumineux les points qu'il vient heurter, créant ainsi une image. C'est un filament similaire qui est utilisé dans un **accélérateur linéaire d'électrons** (linac à électrons pour abréger).

Il existe deux types d'accélérateurs: **linéaires et circulaires**; le CERN possède les uns et les autres. Les accélérateurs utilisent des champs électriques puissants pour injecter de l'énergie dans un faisceau de particules. Des champs magnétiques servent à maintenir une focalisation serrée du faisceau et, dans les machines circulaires, à guider les particules autour de l'anneau.

Un accélérateur **linéaire ou LINAC**, est un accélérateur de particules qui accélère des particules chargées - des électrons, des protons ou des ions lourds - en ligne droite. Des particules chargées entrent sur la gauche et sont accélérées vers le premier tube de glissement par un champ électrique. Dès qu'elles sont parvenues à l'intérieur du tube de glissement, elles se trouvent protégées de l'action du champ électrique et elles dérivent alors à vitesse constante. Quand elles arrivent au prochain intervalle entre les tubes, le champ électrique les accélère à nouveau

jusqu'à ce qu'elles atteignent le prochain tube de glissement. Ceci continue, les particules acquérant de plus en plus d'énergie à chaque intervalle entre tubes, jusqu'à ce qu'elles débouchent de l'accélérateur sur la droite.



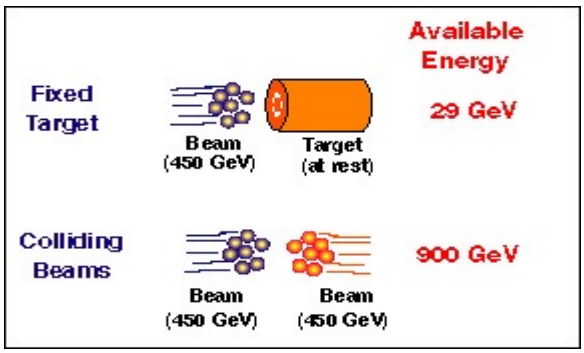
Les tubes de glissement sont nécessaires car un champ alternatif est utilisé et sans eux, le champ alternativement accélèrerait et freinerait les particules. Les tubes de glissement blindent les particules pour le temps durant lequel le champ les freinerait.

Tous les faisceaux de particules du CERN naissent dans des linacs, mais des accélérateurs extrêmement longs seraient nécessaires pour atteindre les énergies dont les physiciens ont besoin. Pour cette raison, les grandes machines du CERN sont circulaires. Les **accélérateurs circulaires** font tourner les particules encore et encore dans un cercle, en leur donnant de l'énergie à chaque tour. De puissants aimants maintiennent les particules dans une trajectoire circulaire, tandis que des champs électriques fournissent la puissance accélératrice. Il y a de nombreux types d'accélérateurs circulaires, et depuis le début des années 30, ils ont joué un rôle décisif dans la physique des particules des plus basses aux plus hautes énergies.
Le plus grand accélérateur du CERN, le grand collisionneur électron-positon LEP, a une circonférence de 27 kilomètres, ce qui adoucit les courbures autant que faire se peut.

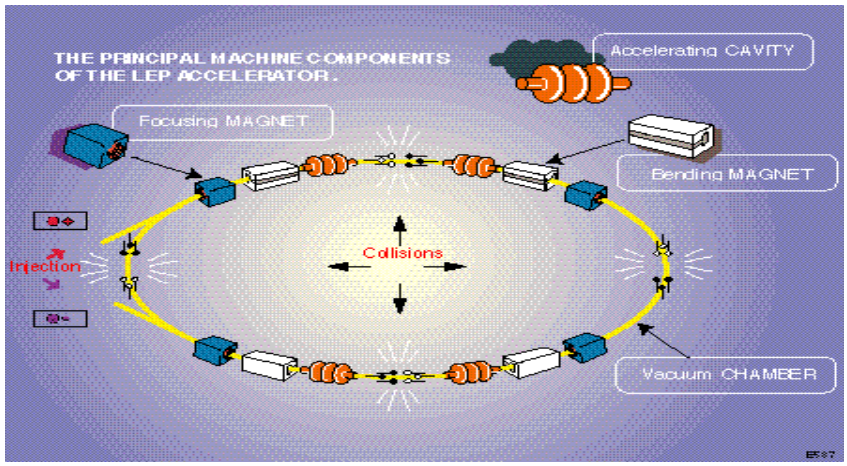
2.2. Le LEP

La célèbre égalité d'Einstein $E=mc^2$ nous dit que matière et énergie sont équivalentes. En conséquence l'énergie d'un faisceau de particules peut être convertie en masse, et donc crée une profusion impressionnante de particules additionnelles, nombre d'entre elles hautement instables et normalement introuvables dans la nature. Cependant si le faisceau incident est simplement

projeté sur une cible fixe, une bonne part de l'énergie du projectile est absorbée par le recul de la cible et n'est donc pas exploitable. Beaucoup plus d'énergie est disponible pour la production de nouvelles particules si deux faisceaux circulant dans des directions opposées entrent en collision. Tous les grands projets mondiaux en physique des particules en préparation sont basés sur de telles machines collisionneuses de faisceaux.



L'accélérateur électron-positon LEP est un collisionneur. Ses 3368 aimants courbent deux faisceaux de particules et les maintiennent en orbite. Là où des électrons chargés négativement sont déviés d'un coté, des positons chargés positivement le seront du coté opposé. Ceci permet au LEP de faire circuler des faisceaux de 90 GeV (*) d'électrons et de positons en sens inverses en utilisant les mêmes aimants. Un autre collisionneur du CERN, le Super Synchrotron à Protons (SPS), utilisait la même technique en accélérant des protons dans un sens et des anti-protons dans le sens opposé.



3. Les divisions du CERN

Pour concevoir et construire l'appareillage sophistiqué du CERN et pour assurer son bon fonctionnement, pour aider à préparer, à mettre en oeuvre les expériences scientifiques complexes, à analyser et interpréter leurs résultats et pour mener à bien la multitude de tâches nécessaires au succès d'une organisation aussi grande et spéciale, le CERN emploie quelques 2500 personnes, hommes et femmes couvrant un large éventail de compétences et de métiers - physiciens, ingénieurs, programmeurs, techniciens, ouvriers qualifiés, administrateurs, secrétaires, ..

Tout le personnel est reparti en divisions qui peuvent être regroupées comme suit:

- **Le secteur de la recherche**

Physique expérimentale **EP**, Physique théorique **TH**, Technologies de l'information **IT**, Education et transfert de technologies **ETT**

- **Le secteur de l'accélérateur**

LHC, **SL**, **PS** direction des accélérateurs **AC**

- **Secteur de la technologie**

Ingénierie et technologies **EST**, Support technique **ST**, Inspection technique et sécurité **TIS**

- **Administration**

Support administratif **AS**, Ressources humaines **HR**, Logistique et fourniture **SPL**, Finance **FI**, services de la direction **DSU**

4. La bibliothèque du CERN

4.1. Organisation du service

4.1.1. Mission

Le service d'information scientifique appartient à la division ETT-SI, et a pour mission la gestion de la bibliothèque et des archives historiques et scientifiques du CERN. A ce titre, il doit acquérir et gérer l'information disponible dans tous les champs scientifiques pouvant intéresser l'organisation et la rendre accessible de la manière la plus simple à la communauté des chercheurs du monde entier dans le domaine de la physique des particules.
Il doit aussi fournir des services permettant d'obtenir l'information qui ne serait pas directement disponible sur le site et tenir les divisions du CERN informées des transformations et des modifications dans l'accès aux documents. Il a aussi pour mission de distribuer les publications du CERN comme les rapports et de sauvegarder les documents et fichiers nécessaires aux chercheurs dans une perspective aussi historique que scientifique.

4.1.2. Le personnel

- **Directeur** : Corrado Pettenati
- **Secrétaire administrative** : Lamia Djerbi

- **Services aux usagers , renseignements bibliographiques , prêts inter bibliothèques**

Chef de section : Jens Vigen

5 autres collaborateurs

- **Gestion des collections : livres, pre-prints et actes de conférences**

Chef de section : Ingrid Geretschläger

7 autres collaborateurs et apprentis

- **Périodiques : revues scientifiques**

Chef de section : Eliane Chaney

un autre collaborateur

- **Archives historiques: CERN et collections Pauli**

Chef de section : Anita Hollier

3 autres collaborateurs

4.2. Le Cern Document Server, service en ligne de la bibliothèque du CERN

4.2.1. Les collections

La liste complète des collections de la bibliothèque du CERN comporte plus de 38 pages, il serait tout a fait inutile de la reproduire ici ou même de l'explorer en détails. Seules les grandes catégories seront présentées avec une attention plus particulière portée a celles sur lesquelles a porté mon travail.

Articles & Preprints (435 944)

Published Articles (141 156) Preprints (239 349) Theses (20 850) Yellow Reports (1056) Reports (26 357) CERN Internal Notes (5028) CERN Committee Documents (2148)

Les preprints sont les articles avant publication soumis directement par les auteurs ou par leur département soit à une archive thématique, soit à une archive institutionnelle. Le CERN en reçoit ou en importe tous les jours dans la mesure ou

les résultats des expériences et les nouvelles pistes de recherche dans le domaine de la physique des hautes énergies ont besoin d'être transmis le plus rapidement possible aux chercheurs. Leur nombre varie grandement d'un jour ou d'une semaine à l'autre, mais à titre indicatif, le 11 septembre 2002 : 254 nouveaux preprints ont été intégrés à la base, 861 depuis le lundi 9 septembre et 682 entre le 2 et le 7 septembre. Les preprints sont des documents clefs dans la perspective de l'auto archivage, c'est donc sur eux que s'est portée mon attention.

Books & Proceedings (48889)
Books (29778) Proceedings (12143) Standards (6968)
Presentations & Talks (14741)
Conference Announcements (12963) Academic Training Lectures (491) Summer Student Lectures (189) General Talks (15) Videotapes (1083)
Periodicals & Progress Reports (2650)
Periodicals (2076) Progress Reports (574)
Multimedia & Outreach (12827) Photos (4084) ATLAS eNews (51) Video Clips (26) Press Cuttings (4399) Exhibition Objects (177) Posters (270) Weekly Bulletin (1,853) HEP Institutes (701) Experiments at CERN (874) Internet Resources (392)
Archives (45656) CERN Archives (42650) Pauli Archives (3006) DSU Archives (0) SL Archives (0)

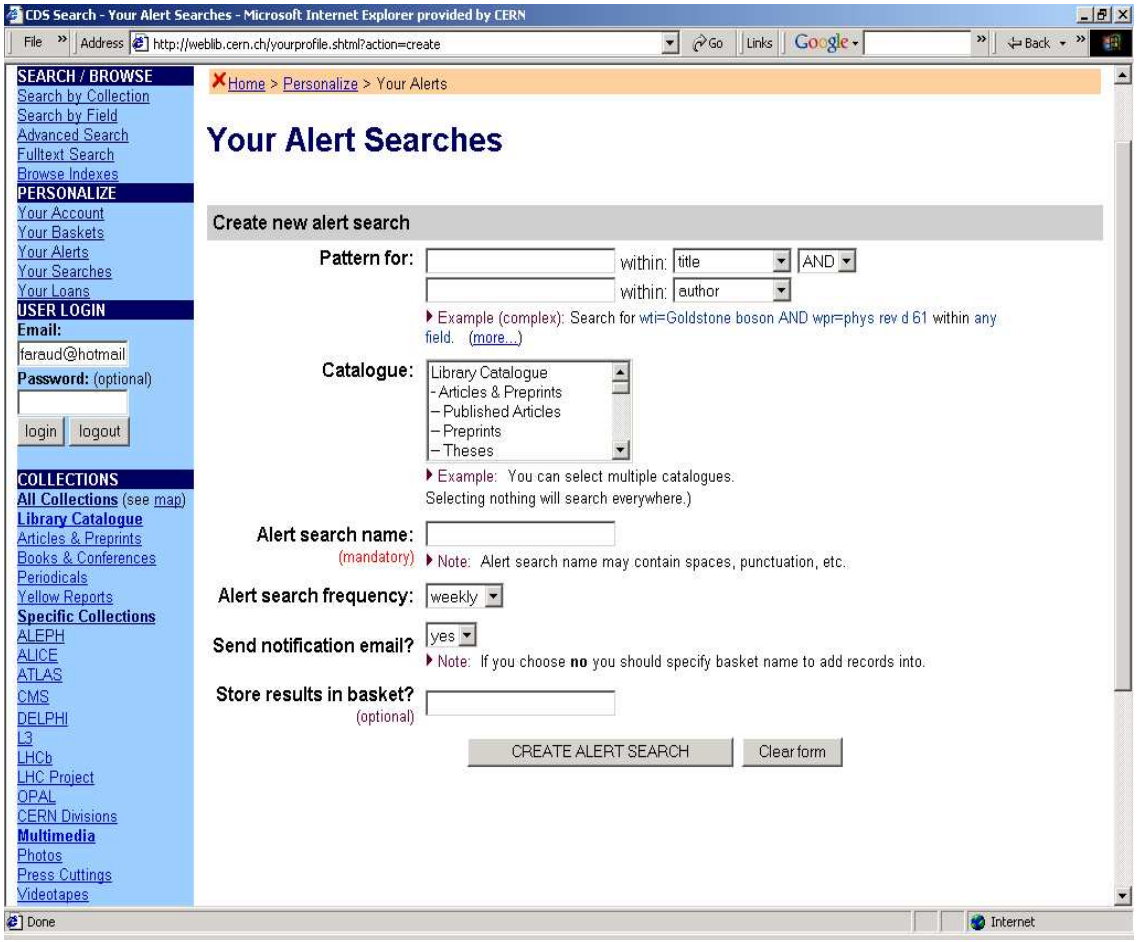
4.2.2. Le moteur de recherche

La bibliothèque du CERN est en train d'opérer une migration de système vers ALEPH 500, pour se faire, une nouvelle version du moteur de recherche CDS a été lancée. Elle repose sur un système MySQL plus rapide, plus souple et plus performant que l'ancien système.
Les recherches peuvent porter sur toutes les collections ou sur une collection particulière ; tous les champs des notices peuvent être interrogés et combinés lors de l'interrogation. Les résultats sont présentés sous forme de liste de notices que l'ont peut ordonner selon des critères à sélectionner (date, numéro de rapport, organisme d'origine…) Les notices des preprints et des articles bénéficient d'un système d'extraction automatique des références bibliographiques: le but étant de

faire un lien vers la version en texte intégral de chaque article ou prepint cité dans la bibliographie d'un document donné.

4.2.3. Les services de recherche

CDS offre un certain nombre de services personnalisés aux lecteurs dont la possibilité de s'enregistrer et de créer des paniers personnels répondant à des requêtes pré-établies. Tout nouveau document entrant dans la base et répondant aux critères choisis sera envoyé dans le panier.



Il est aussi possible d'enregistrer toutes ses recherches, de les sauvegarder et de les représenter ultérieurement au système.

4.2.4. Les services de soumission

La bibliothèque du CERN ayant pour ambition de conserver les documents produits par les scientifiques, les étudiants et le personnel du CERN, CDS est doté d'une interface de soumission des documents, CDS Submit (cf Annexe 1.3)
Selon la collection à laquelle le document appartient, un formulaire est à remplir afin de le soumettre à la bibliothèque. Avant d'être accepté, le document est examiné et commenté par des spécialistes du sujet afin d'établir sa validité ; il est ensuite mis en ligne.

Mission

1. Le contexte

1.1. L'auto archivage

Depuis une dizaine d'années, on assiste à des mouvements de contestation vis-à-vis de la politique des éditeurs commerciaux. La position monopolistique de certains éditeurs, offrant leur contenu à des sommes astronomiques aux bibliothèques, a généré une situation de crise dont de nombreux chercheurs et bibliothécaires ont voulu sortir. La création en 1991, par Paul Ginsparg d'ArXiv, un serveur d'archive dans le domaine de la physique des hautes énergies, destiné à faciliter et à accélérer la communication scientifique a été le début d'une pratique qui devait se développer : l'auto archivage, notion lancée en 1994 par Steven Harnad. Cette pratique a gagné en popularité et touche désormais tous les domaines du savoir. Les chercheurs sont appelés à déposer leurs articles (pré-publications et post-publications) dans des archives électroniques ouvertes afin d'en faciliter la diffusion auprès de la communauté scientifique . L'autre voie qui est explorée est la création de revues alternatives engagées dans le libre accès à des articles acceptés par un comité de pairs.

Afin de fédérer les bases d'archives d'articles scientifiques et de les rendre inter opérables, c'est-à-dire interrogeables simultanément, quelle que soit leur localisation dans le monde, en une seule requête, le mouvement Open Archives Initiative (OAI) fut lancé sur les bases de la convention de Santa Fe tenue en 1999. Ce mouvement définit un ensemble de protocoles techniques liés à la description des données et à leur interrogation, et œuvre pour que la constitution d'archives électroniques devienne un processus établi dans la communication scientifique. Ainsi en 2001, The Open Archives Initiative Protocol for Metadata Harvesting devint la référence : il mentionne les normes qui sont nécessaires pour que les archives deviennent "inter opérables".

1.2. Le CERN et l'auto archivage

Lorsqu'en janvier 2002, le directeur de la bibliothèque du CERN m'a reçue, l'auto archivage était une perspective qui ne tentait que peu de chercheurs et d'institutions. Au début de mon stage, en juin 2002, un nouveau protocole pour la collecte des données venait d'être lancé et le nombre des archives compatibles OAI suivait une courbe exponentielle. De nombreuses initiatives ayant vu le jour en 2002 , l'évolution du CERN, qui travaillait sur un nouveau serveur compatible OAI, dans cette direction était devenue urgente et nécessaire. Ma mission était donc d'étudier le phénomène de l'auto archivage afin de faire une synthèse et de produire des documents destinés à la diffusion interne et à la publication pour motiver des modifications au sein de la communauté scientifique. Pour se faire, il fallait que j'acquière une connaissance approfondie du sujet et que je fasse des propositions pour améliorer les services proposés par CDS dans le cadre de l'auto archivage. Il m'est ensuite apparu au fil de mes avancées que le problème majeur rencontré par l'auto archivage n'était pas d'ordre financier, ni informatique, ni technique mais plutôt, dans bien des cas, d'ordre psychologique, comportemental et sociologique. J'ai fait part de mon intuition à Corrado Pettenati qui m'a alors donne toute latitude pour poursuivre cette intuition et pour produire un article en anglais en vue d'une publication. Cet article n'étant pas termine, il n'a pu être inséré dans ce mémoire mais les grandes lignes en seront données (Annexe 2.4).

2. Méthode de travail

2.1. Phase de familiarisation

2.1.1. Familiarisation avec le domaine

Venant des sciences humaines et de la littérature, j'étais à la fois peu sensible aux problématiques de la physique nucléaire et totalement inculte dans ce domaine. J'ai donc mis à profit, les deux premières semaines de mon stage pour acquérir une culture de base dans ce domaine J'ai donc lu des ouvrages généraux sur la physique nucléaire, sur le CERN et compulsé le site Internet du CERN. J'ai aussi pu assister à certaines conférences générales pour non spécialistes données par Franck Close, éminent physicien anglais, dans le cadre du Programme des étudiants d'été du CERN[1].
Dire que j'étais alors devenue une experte en la matière serait exagéré mais j'ai pu toutefois cerner certaines des problématiques principales de la physique des particules.

2.1.2. Avec le CERN

Le CERN est une organisation colossale qui emploie plus de 2500 permanents et reçoit 8500 physiciens, chercheurs et étudiants du monde entier. L'été est une période ambivalente, entre ébullition et tranquillité: certains sont en vacances alors que d'autres viennent mener des expériences au CERN, ou suivre des cours et assister à des conférences. Toujours lors de mes deux premières semaines de stage, j'ai étudié les organigrammes du CERN, traquer les informations utiles pour me permettre de mieux appréhender cette réalité nouvelle.

J'ai aussi rencontré les membres du service de la bibliothèque et du service informatique responsable du développement de CDS.

2.1.3. Avec la thématique du stage

[1] http://webcast.cern.ch/Projects/WebLectureArchive/lslides2html.php?lid=a01446s4t3

Le sujet de l'auto archivage ne m'était pas tout à fait inconnu dans la mesure ou, lors de ma rencontre avec Corrado Pettenati,en janvier 2002, il m'avait soumis un certain nombre de documents d'actualité afin que je me familiarise avec les grands enjeux.

Toutefois, il fut nécessaire de poursuivre cette première approche et d'engranger de nouvelles informations. Mon deuxième superviseur de stage, Jens Vigen, revenait d'une conférence à Malaga et me confia la lecture des compte-rendus afin que j'entre dans le vif du sujet[2]. Mes lectures m'ont permis de découvrir des sites, des auteurs et des projets qui allaient être des éléments récurrents et fondateurs de ma réflexion.

2.2. Organisation de la mission et plan de travail

2.2.1. Des réunions hebdomadaires avec les tuteurs

Mes deux superviseurs ont souhaité que nous ayons de nombreuses entrevues afin de suivre au plus près les évolutions de mon travail. Lors de la première entrevue avec le chef de la bibliothèque, un plan de travail a été établi: les thématiques à aborder, les principales sources d'information à retenir ainsi que les échéances à respecter pour la transmission des rapports. Les réunions avaient en général lieu le mercredi et le vendredi avec Corrado Pettenati et j'y faisais une présentation détaillée de mes avancées, selon le plan de travail proposé, posais toutes les questions nécessaires et fournissais les premiers jets de mes rapports. Corradao Pettenati de son côté, me transmettait les informations qu'il avait reçues pendant la semaine et sur lesquelles il souhaitait que je me penche.

Les rencontres avec Jens Vigen étaient moins formelles et touchaient l'application CDS. Elles se basaient donc sur une interrogation du serveur, sur une analyse des tables des bases de données, et sur une veille documentaire sur les techniques de classement des documents (document ranking) et de récupération de

[2] *Adaptative Hypermedia and Adaptative Web-Based Systems, Second International Conference, AH 2002, Ma laga, Spain, May 2002*. Paul De Bra, Peter Brusilovsky, Ricardo Conejo, eds. New York : Springer, 2002.616p.
Personalization Techniques in Electronic Publishing on the Web : Trends and Perspectives. Stefano Minzano, carlo Tasso, eds. Malaga : Spain, 2002. 165p.

l'information (information retrieval). Jens Vigen me faisait aussi part de ces idées et de pistes à explorer.

J'ai tenu à travailler en étroite collaboration avec les informaticiens de l'équipe : il me semblait en effet inconcevable de ne pas les tenir informés des orientations futures et des souhaits de la direction de la bibliothèque. Il me semblait tout aussi inconcevable de ne pas confronter les nouvelles idées et les propositions de développement avec ce qui était réalisable par les informaticiens et ce qui était en cours ou ce qui existait déjà ; autrement dit de distinguer le théorique du pratique.

2.2.2. Des présentations régulières au service

Dans la même perspective, Corrado Pettenati a souhaité que je fasse des présentations régulières de mon travail au service et que j'assiste aux réunions des superviseurs (chef de groupe) ou de service.

Munie de transparents et de photocopies, j'ai donc exposé mon travail à tous les membres de la bibliothèque afin de les y associer et de recevoir leurs commentaires et de bénéficier de leur collaboration. Cela m'a permis de m'intégrer au service et de situer mon travail dans le cadre plus vaste de l'évolution de la bibliothèque dans son ensemble.

2.3. Analyse de l'existant sur CDS

2.3.1. L'analyses des tables

CDS repose sur une base de type MySQL constituée de 45 bases de données conservant et traitant l'ensemble des données afférentes au service de la bibliothèque : gestion de collections, des utilisateurs, enregistrement des requêtes, récupération de méta données, de notices, extraction automatique des références. Une des équipes présentes à la conférence de Malaga a lancé un moteur de recherche dédié à la physique des hautes énergies, TORII. Torii propose des services personnalisés aux utilisateurs tels des paniers sélectionnant automatiquement les documents pertinents selon une

grille établie par le chercheur. Ce service existe au CERN, ce sont les Baskets, mais il s'agissait d'analyser leur popularité et l'utilisation qui en était faite.
Les tables suivantes ont donc été interrogées :
Query qui regroupe entre autre le texte des requêtes adressées au serveur et la collection sur laquelle elles portent.
Shelf qui renseigne sur les paniers des utilisateurs
Shelf_record qui renseigne sur le contenu des paniers
User qui renseigne sur les utilisateurs enregistrés
User_query_shelf qui renseigne sur les alertes programmées par les utilisateurs
User_shelf qui lie l'id_user à l'id_shelf

Requêtes effectuées entre le 20/06 et le 30/06:

Shelf : 723 rows
SELECTid FROM shelf where public like 'y' = 84
SELECT id FROM shelf where public like 'n' = 639

Shelf_record
SELECT id_shelf, count(*)
FROM shelf_record group by id_shelf
414 paniers avec au moins un element
6885 records
D'autres requetes ont ete effectuees sur les nombres d'element stockes, les resultats sont synthetises dans un tableau :
SELECT COUNT(id_record), id_shelf
FROM shelf_record group by id_shelf

Nb de paniers	Nb de notices	Nb de paniers	Nb de notices	Nb de paniers	Nb de notices	Nb de paniers	Nb de notices
96	1	4	18	1	42	1	117
53	2	4	19	1	44	1	136
33	3	6	20	1	49	1	149

15	4	3	21	1	53	1	153
25	5	3	23	1	56	1	180
14	6	3	25	1	57	1	189
13	7	2	26	1	60	1	201
9	8	3	27	1	60	1	203
7	9	1	28	2	69	1	215
11	10	1	29	2	70	1	300
9	11	1	30	2	74	1	301
6	12	2	31	1	76	1	307
4	13	1	32	1	82	1	1042
5	14	1	33	1	84	1	1087
8	15	1	34	1	97		
5	16	1	38	1	101		
6	17	1	41	1	102		

User

4976 rows

2448 distinct emails

2448 registred users

user_shelf

SELECT distinct id_user

FROM user_shelf

895 shelves 577 distinct users

user_query_shelf

SELECT distinct id

FROM `user_query_shelf`

1859 alertes

1279 utilisateurs distincts

SELECT id_shelf

FROM `user_query_shelf` where frequency like 'day'

526

SELECT id_shelf
FROM `user_query_shelf` where frequency like 'week
814

SELECT id_shelf
FROM `user_query_shelf` where frequency like 'month'
515

Les requêtes ont permis de constater que relativement peu d'utilisateurs avaient des paniers et que peu étaient actifs (d'autres requêtes portant sur la fréquence de consultation ont été effectuées). La majorité des utilisateurs ayant un panier demande des alertes quotidiennes ou hebdomadaires ce qui indique un besoin fréquent d'information. L'alerte par email informant des preprints correspondant au champs de recherche du scientifique est un procédé qui est utilisé depuis plus de 10 ans parmi les physiciens : il est donc normal que CDS propose ce service. Il serait toutefois bon de le mettre plus en valeur afin d'accroître le nombre de paniers et de le rendre au moins équivalent au nombre d'utilisateurs inscrits, c'est à dire 2448. Un document récapitulant les principales analyses et les conclusions a été transmis à Jens Vigen.

2.3.2. L'analyse du moteur de recherche

La nouvelle version de Cern Document Software a été lancée le 15 août 2002 mais il était possible de travailler sur la version en développement bien avant cette date. Mon travail à consister à effectuer des requêtes poussées portant sur tous les champs et sur toutes les collections et à repérer les erreurs les plus fréquentes et les améliorations à apporter. Ceci a été synthétisé dans un document remis à Jens Vigen dont les principaux axes étaient :

L'analyse des requêtes faites par les utilisateurs : degré de complexité, bases interrogées.

Les options de tri des documents à insérer : tri par date, par nombre de déchargement, par nombre de citations.

Analyse des erreurs dans l'extraction automatique des références : un échantillon de preprints a été prélevé dans la catégorie Astrophysics and Astronomy, correspondant aux numéros de rapports débutant par Astrp-Ph, catégorie où de nombreuses erreurs étaient recensées.

L'échantillon contenaient 399 notices datant d'avril 2001, date choisie parce qu'elle n'était pas trop récente et que ,par conséquent, l'extraction des références avait déjà été effectuée[3].

Les références de chaque notices ont été recueillies, chaque erreur était repérée de manière très simple : le lien était absent. L'absence de lien pouvait être due au fait que la référence portait sur un livre dont le texte intégral n'était pas en ligne, ou au fait que la référence était incomplète ou fausse, ou bien encore due à une non-reconnaissance par le système de la syntaxe ou des abréviations employées.

Les deux premiers cas ne peuvent pas trouver de solutions, en revanche la syntaxe des références peut être comprise. De nombreux auteurs, mettent dans la même référence plusieurs titres, d'un même auteur, en les séparant par un point virgule : le résultat est catastrophique car aucune ne peut être reconnue. Il a donc été proposé que le « ; » dans le corps d'une référence soit assimilé à un retour à la ligne et donc à une nouvelle référence dans la mesure où peu de titres contiennent un point virgule.

Lorsque l'abréviation utilisée n'était pas reconnue, celle ci était notée, et cherchée dans Google afin de savoir à quoi elle correspondait. Par exemple, on trouvait fréquemment « PASP » comme abréviation pour *Publications of the Astronomical Society of Pacific* alors que l'abréviation standard est Publ.Astron.Soc.Pac. ; il n'y avait donc aucune chance de réussir à faire un quelconque lien dans ces conditions…

L'abréviation était saisie dans un tableau qui allait servir a enrichir la liste des abréviations déjà utilisées et reconnues par l'Uploader (application de rapatriement et de conversion des méta données) du CERN.

Ensuite, je vérifiais si le CERN disposait d'un abonnement a la dite revue et je notais le lien vers le texte intégral dans le même tableau. Cela a permis de récupérer des centaines de liens dans toutes la base des preprints, dans la mesure où les problèmes

[3] Il faut préciser que l'extraction des références est un processus automatique qui requiert la surveillance d'un informaticien et qui rencontre parfois des problèmes et des pannes. Il faut en général attendre quelques jours voire quelques semaines après l'arrivée du preprint dans la base, pour pouvoir disposer de ce service de manière satisfaisante.

trouves dans Astro-ph se répercutaient dans d'autres domaines utilisant les mêmes revues.

2.4. Veille documentaire

2.4.1. Les listes et newsletters

Lors de ma première entrevue avec Corrado Pettenati, il m'a fourni des pistes de recherches générales : les sites de l'OAI et les sites des principaux forums.
Les listes et les lettres d'informations suivantes étaient systématiquement vérifiées tous les jours en allant directement sur les sites sans avoir recours aux abonnement par email qui m'ont paru encombrants et lourds:

DigLib la liste de l'IFLA dédiée aux évolutions des bibliothèques digitales. Une liste générale, qui se fait parfois l'écho des problématiques de l'auto archivage.
OAI general list est consacrée aux questions pratiques mais non techniques portant sur le protocole OAI. Très active jusqu'en juin 2002, elle est actuellement en perte de vitesse. Mais l'archive a été utile pour comprendre les questions et les problèmes posés par ce nouveau protocole pour la collecte des méta données.
L'OAI-implementers list est une liste technique où sont exposés les problèmes d'ordre informatiques liés à l'implémentation du protocole.
Le september-forum est une liste de discussion débutée en 1998 et portant sur l'auto archivage. C'est la liste la plus complète, la plus vivante et la plus ancienne portant sur le domaine. L'archive a été d'abord consultée afin d'obtenir une vue générale des problématiques portant sur le sujet.
Le FOS Discussion Forum ,le forum associé à la newsletter FOS de Peter Suber, a été créé en 2001. Il porte sur le Free Online Scholarship, terme lancé par Peter Suber regroupant les initiatives visant à libérer la diffusion scientifique du joug des éditeurs.

La lettre d'information de Peter Suber FOS Newsletter est mise à jour quotidiennement et est un instrument exceptionnel pour la veille. Elle fournit des indications

bibliographiques, des informations factuelles et des listes de liens pertinents. Celle maintenue par Charles W. Bailey, Jr., WebLog, est une bibliographie quotidienne : c'est la aussi un élément indispensable à la recherche documentaire.
SPARC e-news, lettre d'information de SPARC une coalition d'universités, de bibliothèques de recherche et d'organisations engagée dans une action constructive en vue de libérer le système de communication scientifique.

2.4.2. Les principales sources publiées et les sites

Des **introductions** en français à la thématique de l'auto archivage se trouvent sur le site d'Helene Bosc à l'INRA et sur celui de l'ENSSIB.
http://www.tours.inra.fr/tours/doc/comsci.htm
http://www.tours.inra.fr/tours/doc/comscith.htm
http://www.enssib.fr/bibliotheque/ecodoc/Articles/fondateurs.html

- Les sites principaux concernant l'auto archivage

Openarchives Initiative http://www.openarchives.org fournit des informations sur les évolutions du projet.
Open archive forum http://www.oaforum.org site sur l'OAI en Europe, liens sur les conférences et les initiatives européennes.
Eprints.org http://www.eprints.rog site maintenu par Steven Harnad, une des créateurs de l'auto archivage, propose une FAQ très complète et des news.
Budapest Open Archive Initiative : http://www.soros.org/openaccess/ sit de l'initiative de Budapest qui a pour but de favoriser l'accès ouvert aux articles scientifiques publiés et non publies en soutenant l'auto archivage et les revues alternatives ; propose une FAQ très complète et des news récentes.

- Les publications

ARL: A Bimonthly Report on Research Library Issues and Actions From ARL, CNI, and SPARC: Bimensuel de l'association américaine des bibliothèques de recherche.

Ariadne Trimestriel anglais de bibliothéconomie

D-Lib Magazine. Mensuel américain sur les innovations et les évolutions dans le monde des bibliothèques numériques.

Learned Publishing Trimestriel américain de l' Association of Learned and Professional Society Publishers

Journal of Electronic Publishing Trimestriel américain consacré aux méthodes de publications et de diffusions électroniques.

Journal of Digital Information. Revue en ligne anglaise sur le management, la présentation et l'utilisation de documents numériques.

portal. Revue américaine conçue et gérée par des bibliothécaires destinée à couvrir les problématiques actuelles du monde de l'information et de la recherche.

Des articles parus dans Nature et dans Science ont donné lieu à des discussions assez pointues sur le sujet en 2001. Elles ont été consultées :

discussion dans Nature.

discussion dans Science. Divisée en quatre sections.

2.5. Rencontres, enquête et interview

2.5.1. Informaticiens

Apres chaque rencontre avec mes superviseurs, je rendais visite aux informaticiens afin de les tenir informés des avancées et des propositions mais aussi afin de recueillir leurs propositions et de découvrir leurs préoccupations. Jean Yves Le Meur, Tibor Simko et Martin Vesely ont été d'excellents appuis dans mon travail. Leur

parfait connaissance de CDS m'a fait gagner un temps précieux lors des requêtes et de l'analyse des tables.

La rencontre avec Eric van Herwijnen a été motivée par la recherche d'une simplification de l'interface de soumission des preprints. Je pensais qu'un des obstacles à la soumission de preprints aux serveurs d'archives était la complexité des interfaces de soumission. Cette interface comprend plusieurs formulaires à remplir, ce qui rend longue et fastidieuse la soumission. De fait, peu d'auteurs soumettent au CERN même s'il en font partie ; ils préfèrent utiliser la soumission en une seule étape du serveur ArXiv, soumission simplissime mais qui a le désavantage de ne pas permettre une véritable recherche sur les méta données. L'idée d'une feuille de style simple basée sur XML proposée dans un article paru en mars 2000[4] me semblait bonne, j'ai donc voulu savoir ce qu'il en était en 2002 dans la perspective de l'auto archivage. Les chercheurs utilisent le format TEX qui permet d'intégrer directement des équations et des formules Les feuilles de style pour ces documents seraient lourdes et difficilement gérables par les browsers, d'autres part les scientifiques sont très réticents a produire des documents pré formatés.

La réponse de Michel Goosens n'a pas été plus encourageante, XML peut bien être très structuré et fort utile, il ne peut pas faire de miracle lorsque personne ne respecte de structure préétablie. Comment reconnaître un titre, une mention d'autorité, ou un résumé lorsqu'il y a autant de cas différents que d'articles ?C'est impossible ! Il n'est pour l'instant pas possible de faire une interface de soumission en une étape, dans laquelle les meta données seraient extraites automatiquement à partir de documents basés sur une feuille de style unique. Dans les grands instituts, les secrétaires de division devront continuer à remplir de longs formulaires, dans les petits laboratoires, les chercheurs souhaitant déposer leurs articles auront à le faire eux-mêmes.

[4] Eric van Herwijnen. The Impact of XML on Library Procedures and Services. Hep Libraries Webzine, 1, mars 2000

2.5.2. Professionnels de l'information

Les interactions avec le personnel de la bibliothèque ont été incessantes notamment avec les personnes responsables des preprints auxquelles je demandais régulièrement des statistiques et des éclaircissements sur les bases.
Catherine Cart et Jocelyne Jerdelet m'ont fourni des renseignements sur la provenance des preprints et sur l'évolution des modes de soumission.

Un des problèmes soulevés par l'auto archivage est la présence du numéro de rapport dans les notices des documents. Les articles, les preprints, les thèses, les rapports… ont un numéro de rapport. La structure du numéro est la suivante :
Institut d'origine, année, numéro attribué par ordre chronologique.

Par exemple : CERN-TH-2002-209.

Pour le 209eme article de l'année 2002 de la division Théorie du CERN.

Toutefois, un grand nombre de documents, principalement des preprints, ont un autre numéro par exemple : hep-ph/0209103 pour un article en High Energy Physics.Ce numéro provient du serveur d'archive, ArXiv, de Paul Ginsparg, principalement consacre a la physique, toutefois hormis la matiere, il ne donne aucun renseignement : c'est juste un numéro d'enregistrement.
Une de mes propositions était d'analyser les procédures d'attribution des numéros de rapport dans les principaux instituts de physique afin de tenter de les normaliser.
J'ai donc réalisé une enquête (cf Annexe 2.3) distribuée à toutes les divisions du CERN et aux bibliothécaires des instituts dont les noms suivent :

CERN	Toutes les divisions
DESY	H. Preissner
SLAC	Patricia Kreitz
FERMILAB	Heath O'Connell , Robert Atkinson
BNL	Arnold Peskin D.I. Lowenstein Ms. Michiko Tanaka
ANL	Sandy Klepec
CEA	Dr. Joel Feltesse,P. Vernin, Marc Gingold
LBNL	Robert N. Cahn Gail Harper
LAPP	Michel Yvert
KEK	K. Ishii

Jussieu (PAR-LPTHE)	B. Machet
Oxford (OUTP).	R.H. Dalitz

L'enquête fut un succès puisque elle a obtenu un taux de réponses de plus de 90% et a permis de prendre des contacts avec des interlocuteurs qui m'ont épaulée pendant tout mon stage.

D'autre part, j'ai entretenu une correspondance suivie avec Peter Suber le fondateur et l'administrateur du site Free Online Scholarship, faisant état de mes avancées et de mes interrogations.

Propositions d'amélioration

1. CDS et l'auto archivage.

1.1. Self-archiving, super archive, institutional repository, trusted digital repository : des notions à élucider.

1.1.1. Self archiving, auto-archivage

C'est une pratique visant à déposer sur un site personnel, un site institutionnel ou un site thématique le fruit d'un travail scientifique sans faire reposer sa diffusion sur la seule publication. Il peut s'agir d'une pré publication, d'un article publié, ou des résultats d'une expérience, d'un rapport interne, de notes de conférence…
Lorsqu'il s'agit d'un auteur qui dépose sur son site personnel, ou sur un site thématique comme ArXiv, CogrPrint, RePec, NTRS il s'agit **d'auto archivage personnel.**
Lorsqu'une institution adopte une démarche de diffusion en dehors des circuits de la publication des travaux de ses chercheurs, de ses étudiants ou de son personnel, alors il s'agit d'**auto archivage institutionnel**[5].En pratique, dans les deux cas, il faut copier les méta données du document et joindre le texte intégral sur le site afin qu'il soit diffusé gratuitement à toute personne intéressée.

[5] Il s'agit ici d'une présentation simplifiée de la démarche de l'auto archivage. Pour des renseignements plus approfondis, on pourra se reporter à la FAQ de la Budapest initiative http://www.soros.org/openaccess/fr/faq.shtml#self-archiving
ou de l'OAI http://www.ecs.soton.ac.uk/~harnad/selfaq.htm

L'idée s'est développée depuis quelques années et est maintenant adoptée par de plus en plus de chercheurs et d'institutions. Il serait vain de dresser ici une liste des serveurs d'auto archivage qui existent dans le monde car le mouvement a connu une très nette évolution ces derniers mois et on a pu voir les initiatives se multiplier dans tous les domaines [6]. Il existe désormais des *data providers* qui administrent des systèmes compatibles avec le protocole OAI afin d'exposer les méta données relatives à leurs documents. Les *services providers* adressent des requêtes aux systèmes des *data poviders* afin de récupérer des méta données pour fournir un service à leurs utilisateurs.

1.1.2. Superarchive et institutional repository

Partant du principe que le pouvoir de la science réside dans le partage des idées et non dans leur rétention, certaines institutions ont décidé de mettre à la disposition de leurs utilisateurs les notes de travail, les travaux préliminaires, les comptes rendus d'expériences, les notes de cours, le matériel pédagogique des enseignants et des chercheurs , les mémoires de élèves d'une université ou d'un laboratoire[7]. En rassemblant un maximum de matériaux de recherche, les institutions espèrent favoriser la communication entre les chercheurs et aider la recherche. Cette notion de superachive recouvre en fait la même réalité que les *institutionnal repository* tel qu'il est défini par SPARC dans *The Case for Institutionnal Repositories : A SPARC Position Paper :*

> *« digital collections capturing and preserving the intellectual output of a single or multi-university community »* [8]

---

[6] On pourra trouver une liste des fournisseurs de données à cette adresse http://www.openarchives.org/Register/BrowseSites.pl et la liste des fournisseurs de services à cette adresse : http://www.openarchives.org/service/listproviders.html

[7] Jeffrey R.Young . *Superarchives Could Hold All Scholarly Output* The Chronicle of Higher Education, July 5, 2002 disponible http://chronicle.com/free/v48/i43/43a02901.htm

[8] http://www.arl.org/sparc/IR/ir.html page 4.

Il existe encore assez peu de super archives ou *institutionnal repositories* et il est intéressant de se pencher sur leur fonctionnement.

Depuis avril 2002, le **California Digital Library eScholarship Repository**[9], comprenant les départements de sciences humaines des universités Berkeley et d'UCLA, propose aux professeurs de déposer leurs documents de travail mais aussi leurs articles certifies par des pairs- peer reviewed- et soumis à des revues. La transmission se fait par email: le document doit être dans un format accepté PDF, Word, RTF et envoyé avec le titre, l'auteur, le résumé, les mots clefs, le numéro JEL, l'affiliation des auteurs et leurs emails[10]

Les documents sont classés par groupes de travail et accessibles directement[11]:

Working Paper Collections for Affiliated Groups:

- Berkeley Program on Housing and Urban Policy
- Center for Health Research
- Center for International and Development Economics Research
- Competition Policy Center
- Economics Department Working Papers
- Fisher Center for Real Estate & Urban Economics
- Research Program in Finance Working Papers

Le projet le plus ambitieux et qui attire le plus d'attention est celui du MIT, **DSPACE.**

Il prévoit l'ouverture à tous les professeurs de l'institut, et une interface simplifiée de soumission. La seule règle qui prévale à la soumission est que le document doit être complet et dans certains cas, vérifié par un comité dans le département. Les professeurs pourront sélectionner des niveaux d'accès à leurs documents, en en réservant certains aux seuls membres du MIT par exemple. De plus, le projet prévoit des procédures de sélection des documents selon les collections et un processus de validation et de

---

[9] http://escholarship.cdlib.org/repository.pdf et http://escholarship.cdlib.org/

[10] Pour la procédure complète, voir http://repositories.cdlib.org/iber/submission_info.html

[11] **Ohio State Knowledge Bank** fonctionnera sur le même principe, on peut dores et déjà voir le projet final à l'adresse suivante

http://www.lib.ohio-state.edu/Lib_Info/scholarcom/KBproposal.html

vérification à plusieurs niveaux ; un enregistrement des utilisateurs, ces derniers peuvent choisir la collection dans laquelle ils veulent déposer leur documents et doivent fournir de nombreuses méta données. Il sera intéressant de voir comment Dspace aura réussi à faire une interface de soumission simple étant donnes le nombre très important de méta données qui sont requises au moment de l'enregistrement[12].
En Europe, d'ici fin septembre 2002, le projet ARNO academic research of netherlands Online, devrait etre opérationnel, http://cf.uba.uva.nl/en/projects/arno/, ainsi que le serveur de l'universite d'Utrecht, http://www.library.uu.nl/182main.html.

1.1.3. Trusted digital repository

En mai 2002. le Research Libraries Group et le Online Computer Library Center, ont publié un rapport *Trusted Digital Repositories : Attributes and Responsabilities*[13] dans lequel ils définissent les pré-requis et les attributions de ces futures institutions chargées de conserver les documents électroniques. Leur but sera de fournir à la communauté scientifique un accès fiable et sur le long-terme aux ressources digitales dont elle pourrait avoir besoin.
Il est nécessaire d'établir des règles selon lesquelles les institutions seront choisies afin de ne pas confier une telle mission à des institutions qui n'auraient ni les moyens ni les capacités de l'assumer.

Les attributs doivent être les suivants :

Compatibilité avec l'OAIS, Reference Model for an Open Archival Information Service[14] qui fournit une terminologie, des concepts et un modèle fonctionnel.

---

[12] Pour plus de renseignements sur le projet, on peut se reporter à Margaret Branschofsky, Micahel Bass. *Dspace, A Sustainable Solution for Institutional Digital Asset Services* .http://web.mit.edu/dspace/live/implementation/design_documents/functionality.pdf
[13] http://www.rlg.org/longterm/repositories.pdf
[14] DRAFT RECOMMENDATION FOR SPACE DATA SYSTEM STANDARDSReference Model for anOpen Archival InformationSystem (OAIS) http://www.ccsds.org/RP9905/650x0r1.pdf 1999.

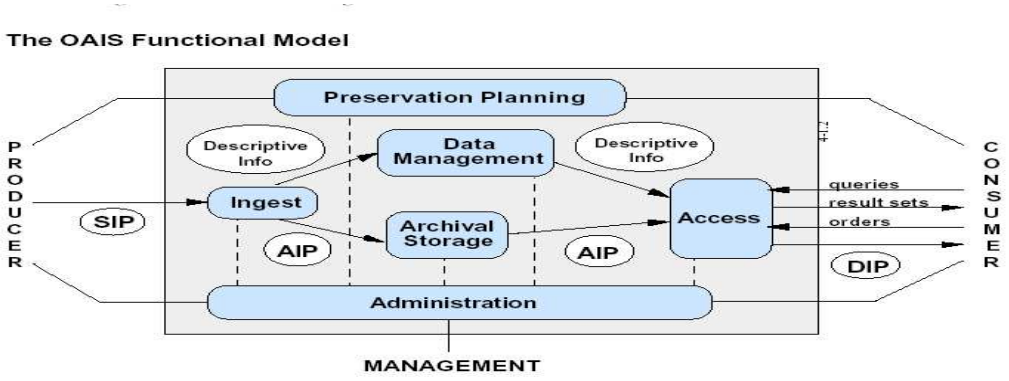


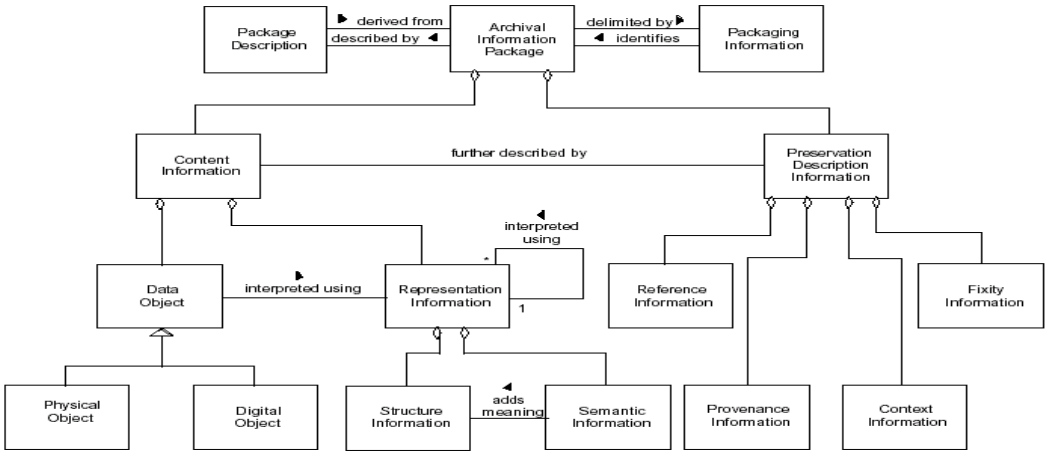
Figure 2. Overview of the Open Archival Information Systems Reference Model

Une organisation administrative solide et dédiée au bon fonctionnement de l 'archive avec l'affectation du personnel compétent pour la gérer et assurer la sauvegarde des données.

Etre une institution de renom et avoir une assise financière suffisante pour persister à long terme.

Fournir une documentation sur son mode de fonctionnement et assurer un accès direct, simple et gratuit à tous.

Garantir la sécurité et la pérennité des sauvegardes en portant une attention toute particulière aux formats de sauvegarde et en s'assurant de la validité technologique de ses choix.



[15]

[15] DRAFT RECOMMENDATION FOR SPACE DATA SYSTEM STANDARDSReference Model for anOpen Archival InformationSystem (OAIS) http://www.ccsds.org/RP9905/650x0r1.pdf 1999.

L'idée est excellente dans la perspective de la conservation des données car c'est un problème majeur auxquels les documents électroniques sont de plus en plus souvent confrontés.

1.2. CDS : vers des activités de plus en plus diversifiées.

1.2.1. Le CERN comme Trusted digital repository

Bien souvent, les documents placés sur des sites personnels disparaissent avec le site lorsque l'auteur n'a plus le temps de l'administrer ; il arrive aussi que certaines institutions ne puissent pas faire face aux pré-requis techniques de la conservation à long terme et , devant faire appel à une société extérieure, préfèrent abandonner toute idée de conservation. Enfin, s'il est nécessaire d'accéder aux documents depuis plusieurs miroirs, il n'est certainement pas utile que toutes les institutions ayant un serveur d'auto archivage et de preprints se lancent dans la conservation exhaustive et à long terme. Dans cette perspective, il serait judicieux d'instituer des organismes chargés d'assurer la conservation dans chaque discipline. Le travail peut être partagé entre plusieurs grands laboratoires en attribuant une tâche à chacun, dans son domaine de prédilection, ou en instituant un organisme par continent par exemple. Le CERN pourrait jouer un rôle majeur à ce niveau dans la mesure où il offre la majeure partie des garanties exigées par le RLG et est une institution mondialement reconnue.

1.2.2. CDS une interface pour un service riche

Le développement de l'auto archivage, soit sur des sites thématiques, soit sur des sites web personnels, soit sur des serveurs institutionnels, ainsi que l'application par des membres de plus en nombreux du protocole OAI sur le transfert des méta données, offrent des nombreuses possibilités au CERN. Il est peu probable de

révolutionner les pratiques des chercheurs qui préfèrent utiliser le service d'ArXiv pour déposer leurs documents et de les voir adopter massivement la soumission au CERN. Toutefois, il est tout à fait possible d'espérer attirer les soumissions de documents internes dans une perspective de super archive visant à regrouper les documents réalisés par les personnels, les chercheurs et les étudiants du CERN. Cette tâche est dores et déjà menée via CDS Submit et régie par la Operational Circular No6. Le problème est que l'ensemble des documents produits par le CERN ne sont pas soumis par ce biais : il faut donc faire une campagne de sensibilisation soit par l'entremise de la bibliothèque soit par celui du Bulletin du CERN en exposant la nécessité de développer ce type de diffusion et d'archivage. Il s'agirait bien entendu de gérer les documents non confidentiels produits par le CERN à l'exclusion des rapports stratégiques, économiques ou d'ingénierie comme ceux stockés dans la bas EDMS relative aux projets des ingénieurs par exemple.

Il est aussi raisonnable de penser que, le nombre d'archives en ligne augmentant, CDS ira sans cesse chercher à rapatrier des données venant de nouveaux serveurs dans le monde entier. Il sera pour ce faire nécessaire de faire une veille scientifique pour débusquer les serveurs qui se développeront dans le domaine de la physique des hautes énergies. On pourra aussi consulter régulièrement la liste des *data providers* fournie sur le site officiel de l'OAI en espérant que les serveurs se déclareront en bonne et due forme et que le site OAI sera maintenu à jour.

Ces évolutions prévisibles entraîneront , elles aussi, une explosion du nombre de documents disponibles et appelleront à de nécessaires modifications de l'interface de CDS afin de proposer des services efficaces et répondant aux besoins des utilisateurs. C'est dans cette perspective que la valorisation des documents en ligne est étudiée.

2. Valorisation des documents en ligne

Le nombre de documents en ligne suit une courbe exponentielle depuis quelques années, processus qui devrait aller en s'amplifiant avec le développement de la pratique de l'auto archivage. Toutefois, cet accroissement quantitatif fait redouter aux

chercheurs un amoindrissement de la qualité ou un manque de signe distinctif faisant perdre tout le bénéfice de cette diffusion accrue. En effet, comment reconnaître un bon document d'un mauvais ? Comment s'y retrouver dans cette masse d'information et éviter le bruit lors des requêtes ? Comment valoriser le document en ligne ?

Il est difficile de répondre à la première de ces interrogations : en effet, la lecture, l'analyse ou la connaissance préalable de l'auteur du document peuvent seules guider le lecteur dans l'appréciation du texte et tous les trésors d'ingéniosité des bibliothécaires et des informaticiens ne parviendront jamais à débusquer les mauvais documents. En revanche, ces mêmes bibliothécaires et informaticiens peuvent aider le lecteur dans sa recherche en ajoutant une « valeur » au document en ligne, valeur immédiatement visible lors de la requête.

Il s'agit donc de trouver des éléments sur lesquels peut s'appuyer cette « valorisation » : une échelle de vérification, le calcul du nombre de déchargements, du nombre de fois que le document est cité ainsi qu'un accès aux commentaires internes semblent être des instruments utiles dans les nouvelles perspectives de développement du service de documentation scientifique du CERN.

2.1. De la vérification à la certification

Le serveur en ligne de la bibliothèque du CERN comprend environ **571 000** références, dont la majeure partie est accessible en ligne, parmi lesquelles on trouve 141 000 articles publiés, les 239 000 preprints, 20 800 thèses, 1060 yellows reports, 26 500 rapports, et 5000 notes internes. Les articles publiés sont accessibles en full text quand le CERN détient un abonnement à la revue qui les a publies. Les preprints sont soit pour la majeure partie rapatriés du serveur ArXiv , soit pour les preprints internes directement soumis au CERN, soit soumis sous forme papier.

L'étude qui suit porte sur les « preprints » présents dans les collections du CERN dont la totalité a soit un numéro de rapport soit un numéro de Los Alamos.

Il est nécessaire de faire état de la confusion qui règne dans la terminologie : un preprint est censé être une pré publication, c'est à dire un document soumis pour

publication, accepté par une revue et en attente de publication. Dans la pratique, l'appellation regroupe tout article qu'il soit destiné à publication ou non. Un e-print, document électronique[16], est un document qui soit n'a pas forcement pour vocation d'être publie, comme les thèses, les rapports, les rapports internes, soit a été refusé par les revue même s'il a été soumis pour publication. Donc, de facto, l'étude porte sur les documents appelés preprints dans le monde de la physique des hautes énergies, c 'est à dire des pré publications auxquelles s'ajoutent des documents qui n'ont pas pour vocation d'être publiés.

Environ 131 000 preprints ont un numéro de Los Alamos soit 51% de la collection.

Environ 134 100 prepints ont un numéro de rapport

Environ 32 700 preprints ont un numéro de rapport délivré par l'institution d'origine et un numéro Los Alamos

Environ 19 160 proviennent des principaux laboratoires : CERN, DESY, FERMILAB, KEK, SLAC, BNL, LBL, ANL, CEA, LAPP, IPTHE.

Le numéro de rapport n'est pas une donnée réellement exploitée dans les serveurs de documents électroniques, pourtant c'est une donnée qui peut s'avérer très riche en informations. L'attribution d'un numéro de rapport recouvre des réalités fort diverses : de l'attribution directe par la secrétaire de division avant même que le papier ne soit terminé, à l'attribution après de nombreuses étapes de relectures, de validation et de corrections. Afin de mieux cerner ces différentes pratiques, une petite enquête a été réalisée avec les principaux laboratoires de physique des hautes énergies dans le monde.

Pour se faire, le questionnaire ci-dessous a été envoyé (voir annexe 2.3 pour la version complète)

Le document est-il vérifié?

[16] Ici le terme document électronique est employé car il reste assez vague et permet d'englober les rapports, les thèses, les conférences… Il a été préféré à e-publication ou à publication électronique dans la mesure ou il serait fâcheux d'assimiler la mise en ligne à une publication.

Le document est-il vérifié par des membres du département, de la division, ou de l'unité?

Le document est-il vérifié par des membres extérieurs?

Le document est-il vérifié par un commission composée de membres du département, de la division, de l'unité et de membres extérieurs?

Le document est-il vérifié par un autre type de commission? Dans ce cas pourriez- vous le décrire?

La diversité des approches est nette, toutefois il est à noter que bien souvent, le document franchit un processus de validation qui comporte plusieurs étapes et engage, outre la responsabilité des auteurs, celle des dirigeants des laboratoires et des unités de recherche.

Il semblerait tout à fait logique que le document porte la trace de ces étapes, ce n'est pourtant pas le cas. En effet, rien ne permet de distinguer un document ayant été vérifié à plusieurs reprises d'un document n'engageant que la responsabilité de l'auteur. On pourrait penser que la présence du Numéro de Rapport (RN) garantisse à elle seule un certain niveau de vérification du document mais la diversité des cas ne permet pas de tirer une telle conclusion et de faire de la présence du RN le corollaire de la vérification et de la certification.

Il semble que l'adjonction d'un champs de méta données et d'un pictogramme (★) seraient des solutions valables.

Organisme	Division	Type de document	Processus de validation
CERN	EP	Preprints	Vérification par les membres de la collaboration et du comité Transmission au secrétariat de la division • LEP transmission au referee compétent pour commentaires et modifications. • non LEP transmission directe des auteurs ou des secrétaires de groupes Puis transmission au secrétariat division et au referee compétent

			Enfin un 'editorial board' valide le document pour certaines expériences
	TH	Preprints	Demande de numéro de rapport par le chercheur Si un des auteurs est anglais, attribution directe si l'auteur est dans la division depuis plus de deux mois Si aucun auteur de langue anglaise, le texte est corrigé avant l'attribution.
	SL , ST , IT		Pas de réponse
	LHC	Rapports de division Rapports projet LHC Notes projet LHC	Vérification et autorisation par chefs de groupes puis transmission au chef de division Vérification et autorisation par chefs de groupes puis transmission au chef de division , autorisation finale donnée par le chef de projet Procédure simplifiée : vérification par chefs de groupes puis chef du projet. Documents sous la responsabilité de l'auteur.
	PS	Tout document	Vérification par chef de groupe et editorial board
	EST	Tout document	Vérification et validation par superviseur, chef de groupe, chef de division
	TIS	Tout document	Vérification par chef de groupe, puis chef de division
	Toutes divisions	CERN OPEN	Documents dont les auteurs ne sont pas intéressés par le processus de vérification ou auteurs non autorisés à présenter leurs documents dans une division du CERN
CEA	DAPNIA	Preprints	Validation par équipe de recherche ou collaboration Soumission par le biais d'une demande d'autorisation de publication Validation par chef de service et par chef du département.
IPTHE	Toutes	Preprints, thèses	Ecrits sous la responsabilité du chercheur Thèses vérifiées par le directeur de thèse
FERMILAB	Expériences Théorie Astrophysique	Preprints	Théorie et astrophysique : pas de processus de validation Autre division : validation par les membres de la collaboration Soumission et vérification de la provenance puis transmission a un spécialiste du sujet Modification si nécessaire et transmission au bureau d'enregistrement des brevets
DESY		Preprints, thèses, rapports internes, compte-rendus, auto-archivage	Rapports internes : validation par le chef de groupe ou le porte-parole de la collaboration Thèses : validation par l'université de l'auteur et par un membre de la collaboration ou par un spécialiste du sujet Rapports DESY : validation par un scientifique spécialiste du sujet et par le directeur de DESY en cas de publication Comptes-rendus : validation par le directeur général E-prints : validation par le directeur général et par un scientifique

			spécialiste du sujet
BNL	accélérateur	Tout document	Validation par le chef de division et par le directeur du département
KEK		Tout Document Progress report, Report, Internal, Proceedings, preprints	Validé par le comité de la bibliothèque et des publications. Ce comité comprend des membres de tous les instituts mais chaque document est vérifié par le représentant de l'institut dont il dépend. Numéro de rapport et valide par le comite KEK Pour les preprints et pour les proceedings vérification supplémentaire par le chef de l'institut concerné. Les Annual reports sont validés par un comité éditorial
Oxford OUTP		Tout document	Validation et acceptation par une revue en cas de publication. Pas d'autre processus de validation
LBL		Tout document	Vérification et validation par un reviewer de la division désigné par le chef de division

La physique des hautes énergies compte une multitude d'équipes et de laboratoires de recherche dans le monde entier, il serait donc vain, voire même impensable, de faire une enquête générale sur les processus de validation des documents dans chacun des ces laboratoires. En revanche, une étude de moyenne ampleur portant sur un échantillon raisonnable de laboratoires et d'équipes devrait permettre d'établir une échelle à l'aune de laquelle pourrait être mesurée la validation du document. Ainsi, au moment de la création de la notice, un champ supplémentaire des méta données serait rempli, renseignant sur le degré de vérification du document. On peut appeler cela comme le fait William Arms[17] une « quality metadata »soit une méta données portant sur la qualité

En se basant sur l'enquête réalisée, on peut déjà établir une échelle de valeur :

0 pour les documents n'engageant que la responsabilité de l'auteur

[17] Arms, William Y. *Quality control in scholarly publishing. What are the alternatives to Peer review ?* http://doc.cern.ch/archive/electronic/other/agenda/a01193/a01193s4t3/transparencies/Arms.ppt donné lors du worskhop OAI du CERN en 2001.

1 pour les documents engageant la responsabilité de la collaboration ou du groupe

2 pour les documents engageant la responsabilité de la collaboration ou du groupe et vérifiés par le chef de division

3 pour les documents engageant la responsabilité de la collaboration ou du groupe et vérifiés par le chef de division et vérifiés par un comité spécifique comme par exemple le comité de la bibliothèque de KEK, ou le comité scientifique de DESY

4 pour ceux soumis à une revue et ayant de fait eu une vérification interne supplémentaire

L'échelle s'applique donc aux documents électroniques émanant d'institutions de recherche et non aux articles publiés ou acceptés par des revues et ayant de fait été soumis au peer review. Le fait que le document ait été accepté ou non par une revue ne devrait pas influer sur le statut du document à ce stade. En effet, si le document est accepté et va être publié un pictogramme pourra le signaler et un autre champ serra ajouté si besoin est, afin de faire le lien avec la version publiée. Mais il semble peu judicieux de faire coexister sur le même plan un processus de validation interne et la sélection opérée par les revues dans le cadre de la publication. En revanche, si un document a été soumis à une revue en vue d'une publication et a de fait subi une vérification interne supplémentaire (voir cas 4 de l'échelle) alors cela devrait être précisé.

Toutefois, il est impératif de faire une échelle simple ne comportant pas plus de 5 niveaux qui pourront être matérialisés par des étoiles ou par un autre pictogramme, directement visibles sur la référence du document. Le but n'est pas d'égarer le lecteur et de satisfaire un fantasme bibliothéconomie visant à englober X niveaux de vérification, mais de lui fournir un outil simple, visible et efficace.

Toutes ces étapes de vérification débouchent sur une certification institutionnelle comme on peut le voir dans le modèle proposé par SPARC :

Scholarly Communication Functions in a New Disaggregated Model [18]

Function	Process	Actors	Process Sponsor
Registration	Posting electronic paper to repository	Academic author-researcher	Repository sponsor
Certification	Peer review Associative certification Online response	Academic referees Academic respondents	Overlay journalsAcademic departmentsRepository sponsor
Awareness	Interoperable open repositories and support services	Librarians	Academic institution Professional Societies Third-party providers
Archiving	Perpetual access	Librarians	Academic institution

2.2. Modifications de l'interface ajout de fonctionnalités

Dans la même perspective de valorisation du document en ligne, un certain nombre de fonctionnalités peuvent être ajoutées et s'avérer très efficaces. Celles-ci peuvent s'appliquer à tous les documents que ce soit de pré publications, des articles publiés, des thèses, des rapports etc… Elles seront ajoutées dans l'interface du CERN Document Server (CDS).

2.2.1. Nombre de déchargements

Le nombre de déchargements d'un document pourra renseigner sur sa popularité, sur l'intérêt qu'il aura suscité dans la communauté scientifique. Toutefois, et cela a été

[18] In *The Case for Institutional Repositories : a SPARC Position Paper*. Prepared by Raym Crow, May 2002 http://www.arl.org/sparc/IR/ir.html

souligné dans la FAQ du serveur ArXiv[19] ce système n'est pas sans inconvénients : un auteur pourrait artificiellement et automatiquement décharger son document afin de faire croître sa popularité ; certains articles ou rapports sont juste consultés rapidement et ne sont pas conservés l'intérêt du lecteur est donc difficile à cerner et il est impossible de savoir si le document est effectivement lu ; un auteur pourrait être très embarrassé de voir le manque d'intérêt que suscite son article ; enfin il serait nécessaire de conserver l'impression de liberté des lecteurs et donc de ne pas enregistrer les opérations faites sur le web. Ces obstacles n'en sont pas vraiment : au sein d'une communauté scientifique sérieuse, il est peu fréquent de voir des pratiques de falsification. D'autre part, le problème de lecture se pose aussi avec les ventes de livres ou les emprunts : combien de livres a-t-on empruntés et ramenés sans les avoir lus ? combien a-t-on achetés et stockés sur des étagères surchargées ? Pourtant, les libraires comptabilisent ces ventes, et les bibliothèques comptabilisent aussi ces emprunts ! Il en va de même pour un document électronique : le déchargement ne signifie pas la lecture, il indique un intérêt pour le document. Quand au problème d'ego malmené par le manque de popularité d'un document, il semble la que ce soit un argument fallacieux. Un chercheur adulte doit être capable de supporter ce genre de contrariétés et d'échecs qui sont riches en enseignements et permettent l'amélioration des prochains papiers. Enfin, *last but not least*, la question de la sauvegarde de la liberté des internautes : il est très clair que des statistiques du type « tel chercheur, de telle université a déchargé tel document » serait une atteinte a la liberté mais comptabiliser de manière anonyme, le nombre de déchargements ne semble pas relever de ce type de violation. Cela peut être comparé aux compteurs sur les sites qui indiquent le nombre d'internautes présents à un moment donné, ou le nombre de requêtes adressées à un système ; compteurs qui sont déjà mis en place sur CDS au CERN. Donc le décompte de l'accès au document peut être un outil parmi d'autres pour évaluer sa popularité, on pourra au besoin l'accompagner d'une mise en garde sur sa fiabilité due aux risques de falsification.

Ce compteur pourra facilement être mis en place au CERN dams la mesure où il existe déjà un outil, Frequency of Fulltext Document Download on CERN Document Server, dans lequel seules les URL sont conservées. Les robots connus sont éliminés et seuls

---

[19] Voir E-print Archive Help Faq : http://arxiv.org/help/faq/statfaq

les documents complètement déchargés sont comptabilisés. Puis, les informations essentielles, c'est à dire la clef du full text dans la base bibliographique de CDS, sont récupérées par le biais de l'uploader du CERN. Après vérification, et élimination de déchargements multiples par le même hôte, les données sont sauvegardées dans Impact.[20]

Pour l'instant, il faut saisir le RN du document recherche et sélectionner une période, mais d'ici peu il sera possible d'intégrer cette fonctionnalité directement dans la notice du document. Toutefois, le comptage en temps réel et directement dans la notice ne pourra être obtenu car cela ralentirait le serveur et la communication des résultats des requêtes ; il suffira de cliquer sur un bouton pour actionner cette fonction.

2.2.2. Nombre de citations

Cela est déjà mis en place au CERN, la procédure directe serait trop coûteuse et ralentirait la communication des résultats mais le système peut le faire sur requête. Il repose sur l'extraction automatique des références dans le corpus du CDS mise en place par l'équipe de Jean-Yves Le Meur. Chaque document est analysé et ses références sont repérées, extraites, stockées, et liées à la version en ligne avec un taux de réussite dépassant 80%. A partir, de là il est possible de considérer que le comptage du nombre de fois qu'un article a été cité est tout à fait réalisable et fiable.. Le nombre de citations reflète l'utilisation faite d'un article par la communauté scientifique ce qui en fait un bon outil d'appréciation des documents.

La présentation de la notice devrait être repensée et rendue un peu plus claire : l'idée de présenter les références directement est excellente, en revanche les outils de traitement du document ne sont pas assez mis en avant.

Les trois lignes techniques spécifiant les formats en haut à gauche devraient être mise au bas de la notice car elles n'intéressent pas forcement les lecteurs.

---

[20] Pour de plus amples renseignements voir http://doc.cern.ch/impact/impact_tech.html et http://doc.cern.ch/cgi-bin/impact/search.pl

On peut s'inspirer de la présentation faite sur Citebase[21], qui tend à regrouper thématiquement les données : les références, les citations… et faire une présentation plus structurée et donc plus efficace.

Un lien vers la version publiée pourra être inséré dans les notices des pré publications lorsque toutes les notices seront intégrées dans CDS afin de laisser au lecteur le choix de la version qu'il veut utiliser.

Il serait utile de spécifier le type de document dont il s'agit : on a vu que le vocabulaire était flou et qu'il y avait une tendance à assimiler preprints et e-documents, pré publications et documents électroniques, ce qui n'est pas la même chose. Il serait donc souhaitable de créer un champs stipulant s'il s'agit d'une pré publication ou bien d'un document non destiné à la publication : rapport interne, scientific notes, document auto-archivé…Il est bien évident qu'un telle distinction doit s'appuyer sur un champs de méta données qui devrait être ajouté dans la notice MARC du document. Toutefois, cette distinction n'a aucune raison d'être si un effort général n'est pas fait en terme de terminologie…

Si l'idée de l'échelle de certification est retenue, on pourrait inclure des étoiles ou tout autre pictogramme, pour signifier le niveau du document.

Les références peuvent être conservées en l'état, en revanche, il serait bon de réorganiser la partie gauche de la notice en groupant les moyens d'accès au document : déchargement, conversion, disponibilité. Puis les outils proposés : nombre de déchargements, nombre de citations, recherche dans le texte intégral, extraction automatique des mot-clefs. Enfin, les services proposés : impression, extraction des données chiffrées, envoi des documents par e-mail et les différents formats de la notice.[22]

2.2.3. Ajout de fonctionnalités dans le cadre de super archive

---

[21] Voir un exemple http://citebase.eprints.org/cgibin/citations?id=oai%3AarXiv%3Aastro%2Dph%2F9712212

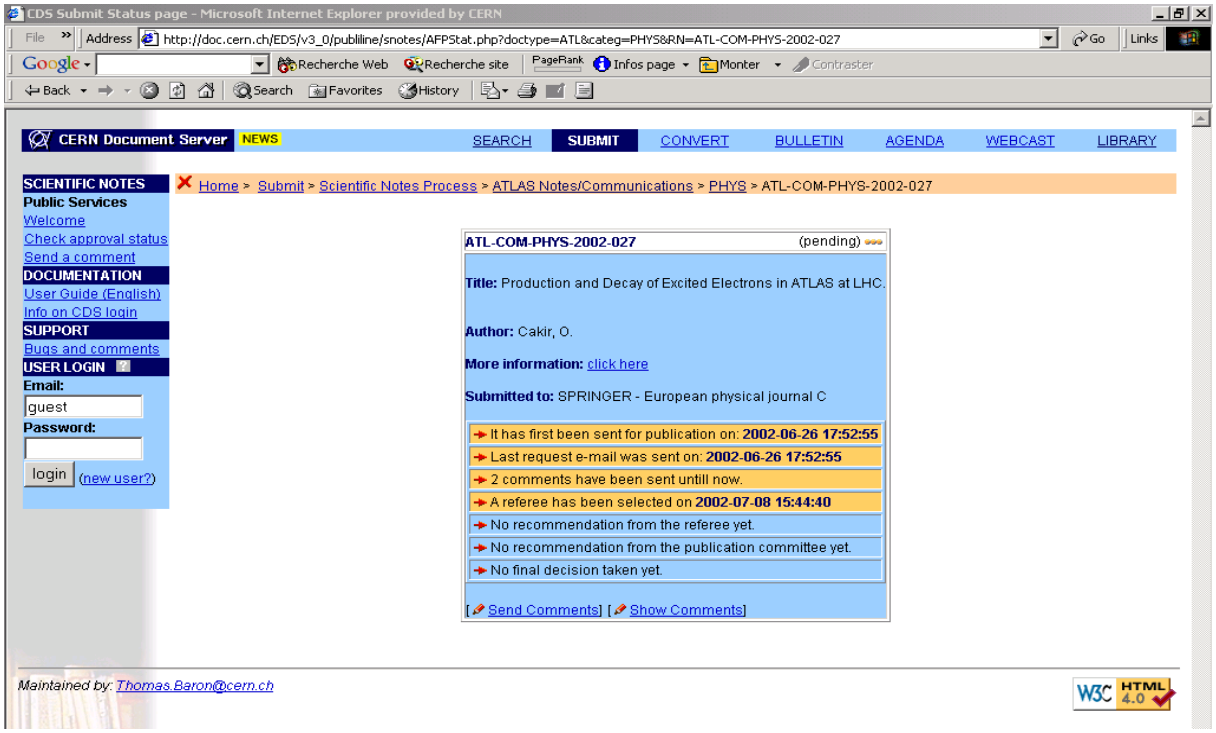
[22] Voir annexe 2.1 pour l'application des modifications à une notice.

CDS va être amené à jouer plusieurs rôles : il sera l'interface d'accès à des documents numériques en physique des hautes énergies conservés sur le long terme , l'interface de dépôt de documents internes du CERN dans la perspective d'auto-archivage institutionnel et une interface de consultation et de recherche en tant que data povider et serveur d'eprints. Si l'optique de conserver tous les documents émanant des chercheurs du CERN est conservée et si les chercheurs adhèrent à cette idée, alors il sera nécessaire, outre le fait d'ajouter des fonctions de valorisation automatique des documents en ligne, d'ajouter aussi des outils de valorisation manuelle et de service personnalisé aux lecteurs. Il serait utile

De faire un accès spécifique pour la super archive : si il se connecte depuis la bibliothèque, l'utilisateur sera reconnu par son mot de passe utilisé pour démarrer la session, il en ira de même depuis son bureau au CERN

D'offrir la possibilité de soumettre son document, et de vérifier son statut comme c'est déjà le cas par CDS Submit

De laisser les commentaires visibles pendant que le document est en cours d'acceptation comme c'est le cas pour les Scientific Notes et de permettre aux usagers de faire des commentaires s'ils le désirent.



Il ne serait pas ici question de transformer ces commentaires en statistiques comme le propose Nathalie Denos dans son article sur Torii[23], en évoquant des grilles de qualité remplies par les lecteurs au fil de leurs lectures… Il est assez difficile de mobiliser les chercheurs pour remplir des champs de méta données portant sur leurs articles, il paraît impensable, pour l'heure, de leur demander de consacrer du temps aux travaux de leurs collègues.

Par contre, l'idée d'opérer un social filtering, c'est à dire un filtre par affinités si on peut dire, est intéressante. Les utilisateurs ont déjà la possibilité d'avoir un panier et un profil sur CDS : il existe 889 paniers appartenant à 572 utilisateurs distincts. La plupart ne contiennent que peu de notices, entre 1 et 5, ce qui prouvent que les paniers ne sont pas utilisés à des fins de stockage ou pas utilisés du tout ! Sur les 2780 utilisateurs qui ont enregistrés un profil sur CDS, 1859 fonctionnent par alertes dans leur email : 526 alertes quotidiennes, 814 hebdomadaires, 515 mensuelles. Ce nombre augmentera avec le nombre grandissant de nouveaux documents présents sur CDS, mais il serait quand même utile de faire un peu de « publicité » autour de ce service…

[23] *Human Evaluation of Documents Benefit to the Community* in The Second International Conference Adaptative Hypermedia and Adapatative Web Based Systems. Malaga Spain, May 2002. Stefano Mizzaro and Carlo Tasso Eds , p 105-114.

Conclusion

Le CERN est au cœur des évolutions et des problématiques actuelles de la diffusion de la documentation scientifique : l'auto archivage institutionnel, personnel, thématique, les archives institutionnelles et les trusted repositories, ainsi que les questions d'archivage à long terme et la migration des formats.
Quelques fonctionnalités vont être ajoutées à CDS mais il semble que beaucoup d'éléments soient d'ores et déjà réunis pour permettre l'intégration de CDS dans un ensemble plus vaste qui serait un véritable fournisseur de données et de services, charges de collecter des données dans le monde entier se rapportant à la physique des hautes énergies.

Toutefois, l'élément essentiel ne réside pas dans les améliorations informatiques et bibliothéconomiques mais dans l'attitude des chercheurs –utilisateurs et lecteurs. Ils sont la clef de voûte de la diffusion scientifique: il est indispensable de convaincre les chercheurs, de faire des enquêtes, des analyses sociologiques pour cibler leurs comportements face à la lecture et à la recherche d'information afin de leur proposer des services réellement adaptés à leurs besoins et non à l'idée que nous pouvons nous en faire.

Bibliographie

A self-archiving thematic bibliography

January- September 2002

Avertissement :

La présente bibliographie se veut aussi complète que possible sans pour autant viser à l'exhaustivité. Elle regroupe les articles et les rapports, publiés ou non, parus entre janvier et septembre 2002, sur le sujet de l'auto-archivage. Les communiqués de presse en sont exclus et ainsi qu les articles des quotidiens d'information générale. De nombreuses présentations données lors de conférences ne figurent pas ici, car elles n'étaient pas encore mises en ligne lors de la réalisation de cette bibliographie[24].

General :

Brody ,Tim. *... and the Bigger Picture* Presented at the *CURL ePrints workshop*, Glasgow. March 2002. (Powerpoint presentation) also available as (Powerpoint) html.

Case , Mary M. and Prudence S. Adler. *Promoting Open Access: Developing New Strategies for Managing Copyright and Intellectual Property* **ARL Bimonthly Report on Research Library Issues and Actions** , No.220, February 2002

D-Lib Editorial . *Digital Preservation – A Long Journey* . **D-Lib** Vol. 8, No. 5, May 2002.

[24] Le FosNews de Peter Suber est un excellent outil de veille pour se tenir informé des nouvelles parutions et des nouvelles mises en ligne sur le sujet.

Eprints Forum. *Self-Archiving FAQ* **University of Southampton.**

Fox ,Peter. *Archiving of electronic publications - some thoughts on cost* **Learned Publishing.** Vol.15, No.1, 2002, p: 3-5.

Greenstein, Daniel, and Suzanne Thorin. *The Digital Library: A Biography*.. Washington, DC: Digital Library Federation and Council on Library and Information Resources, 2002

Hitchcock , Steve The Emerging Framework for Scholarly Communication Presented at *The Future of Journal Publishing*, Nottingham. March 2002. (Powerpoint presentation)

Hitchcock , Steve Revealing a New Dynamic: Interaction in an Open Access Archive, Presented at the First Workshop of the Open Archives Forum, Pisa. May 2002. (Powerpoint presentation)

Hunter, **Philip.** *An e-prints revolution?* **Ariadne,** No. 31, April 2002**.**

Needleman, Mark. *The Open Archives Initiative.* **Serials Review,** Vol.28, No. 2, Summer 2002, p.156-158.

Odlysko, Andrew. *The rapid Evolution of scholarly communication*. **Learned Publishing.** Vol.15, No.1, 2002, p.7-19.

Rauner, Max. *Wissenschaft im Cyberspace: Das Internet als Alternative für Fachzeitschriften* **Neue Zurcher Zeitung,** August 2, 2002.

Suber, Peter. *Budapest Open Access Initiative:Frequently Asked Questions*

Suber, Peter *Where Does the Free Online Scholarship Movement Stand Today*? **ARL Bimonthly Report on Research Library Issues and Actions** No.220, February 2002.

Suber , Peter *The Free Online Scholarship Movement: An Interview with Peter Suber*. **The Technology Source** , September/October 2002.

Suber , Peter . *Momentum for Eprint Archiving* **FOS Newsletter** , August 8, 2002,

Tennant, Roy. *Institutional Repositories*. **Library Journal**, 15 September 2002

Van de Sompel ,Herbert and Carl Lagoze . *Notes from the Interoperability Front: A Progress Report on the Open Archives Initiative* . 2002. 6th European Conference on Research and Advanced Technology for Digital Libraries, Rome,September 2002.

Projects and programs.

A report by Ariadne. *Open Archives Forum - First Workshop: Creating a European Forum on Open Archives. on the workshop which took place 13-14 May 2002, at the CNR, Pisa..* **Ariadne,** No.32. June July 2002.

Carpenter, Leona . *Open Archives Forum - a 'place' for you to meet* Presentation made at the OAF Workshop in Pisa, May 2002. (Powerpoint document)

Carpenter, Leona ; Donatella Castelli and Susanne Dobratz. *Overviews of European Activity* Presentation made at the OAF Workshop in Pisa, May 2002. (Powerpoint document)

Crow, Raym. *The Case for Institutional Repositories: A SPARC Position Paper* . SPARC, 2002.

Dobratz, Susanne Birgit Matthaei and Dr. Peter Schirmbacher *Open Archives Forum (OAF)*. **DLIB**, Vol 8 No.2, February 2002.

Dobratz; Susanne, **Friederike Schimmelpfennig** and **Peter Schirmbacher** *The Open Archives Forum* **Ariadne,** No.31 April 2002

National Library of Australia Electronic Information Resources. *Strategies and Action Plan 2002-2003*:

National Library of Australia and the State Library of Tasmania.*Guidelines for the Creation of Content for Resource Discovery Metadata*:

Poynder, Richard. *George Soros Gives $3 Million to New Open Access Initiative*, **Information Today NewsBreaks**, 18 February 2002.

Pressler, Christopher. *Update on the FAIR Programme*. **D-Lib** , Vol.8, No. 7/8, July/August 2002

Research Libraries Group. *Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report***.** RLG; Mountain View, CA: May 2002

SPARC.*Gaining Independence: A Manual for Planning the Launch of a Nonprofit Electronic Publishing Venture*:

Straccia, Umberto. *CYCLADES - An Open Collaborative Virtual Archive Environment***.** Presentation made at the OAF Workshop in Pisa, May 2002. (Powerpoint document)

Tools and services:

Asnicar, Fabio .*Torii - Access the Digital Research Community* Presentation made at the OAF Workshop in Pisa, May 2002. (Powerpoint document)

Brody ,Tim;Leslie Carr;Stevan Harnad; Xiaoming Liu ;Kurt Maly ;Michael L. Nelson; Mohammad Zubair. *A Scalable Architecture for Harvest-Based Digital Libraries - The ODU/Southampton Experiments***.** Preprint, ArXiv, 2002.

Carr, Leslie .Hypertext, Digital Libraries and Us: Changes in Scholarly Communication. OpCit within the broad context of hypertext and digital library research, past and future. . Presented at Nottingham, July 2002

Geisler , Gary ; Marilyn McClelland; David McArthur and Sarah Giersch .*Challenges for Service Providers When Importing Metadata in Digital Libraries* **D-Lib** , Vol. 8, No. 4, April 2002.

Hagedorn ,Katrine. *Launch of OAIster Project* **D-Lib,** vol 8 No.3, March 2002.

Hagedorn , Kat *OAIster Search Interface Launched* **D-Lib** , Vol.8, No. 7/8, July/August 2002

Hitchcock, Steve. *From eprint archives to open archives and OAI: the Open Citation project* **.** Presentation made at The All Projects Meeting, Edinburgh, 24-25th June 2002

Hitchcock , Steve . *Citebase: an OAI Citation-ranked Search Service*. **D-Lib** , Vol.8, No. 7/8, July/August 2002

Lagoze ,Carl *Core Services in the Architecture of the National Digital Library for Science Education* ArXiv, preprint 0201025, 2002. 2002.

LeFurgy, William G. *Levels of Service for Digital Repositories* **D-Lib** Vol. 8, No. 5, May 2002.

Nicholls ,Bill.. *Open Meta Tools*. **Byte Magazine**, Feb. 25, 2002

Pullinger ,David. *Instant linking - delayed use: setting provider expectations* **Learned Publishing**. Vol.15, No.1, 2002,p.21-25.

Vesely Martin *CERN Dokument Server Software* Presentation made at the OAF Workshop in Pisa, May 2002. (Powerpoint document)

Peer review

Arms, William Y. *What Are the Alternatives to Peer Review? Quality Control in Scholarly Publishing On The Web*. **The Journal of Electronic Publishing** , No.8, August 2002.

Rowland, Fitton. *The Peer-Review Process*. **Learned Publishing,** Vol 15, No 4, October 2002, p 247-ff

Spigler, Renato. *Peer Reviewing and Electronic Publishing* **CERN Webzine**, No.6, March 2002.

Metadata and Protocol

Breeding, Marshall. *Understanding the Protocol for Metadata Harvesting of the Open Archives Initiative*. **Computers in libraries**, September 2002.

Brody ,Tim. OAI Protocol for Metadata Harvesting, Presented at Applications of Metadata, a one-day conference organised by the BCS Electronic Publishing Specialist Group, London. May 2002. (Powerpoint presentation)

Cole, Timothy W. Qualified Dublin Core metadata for online journal articles **OCLC Systems & Services,** Vol.18, No. 2002: p.. 79-87.

Duval, Erik ; Wayne Hodgins; Stuart Sutton ;Stuart L. Weibel. *Metadata Principles and Practicalities* **D-Lib** Vol. 8, No. 4, April 2002.

Heery, Rachel and Harry Wagner. *A Metadata Registry for the Semantic Web* **D-Lib** Vol. 8, No. 5, May 2002.

Hitchcock , Steve A brief overview of the Open Archives Initiative and OpenURL Presented at the *JISC All-Projects Synthesis Meeting*, Manchester. January 2002. (Powerpoint presentation) also available as (Powerpoint) html

Kaczmarek , Joanne. *OAI-PMH Project at the University of Illinois at Urbana Champaign* **D-Lib** Vol. 8, No. 4, April 2002.

Johnston, Peter. *Metadata Sharing and XML.* **Nof-digitise** , Technical Advisory Service Bath, UK: UKOLN, 2002.

Nelson ,Michael. *New Developments in OAI (OAI-PMH 2.0)* Presentation made at the OAF Workshop in Pisa, May 2002. (Powerpoint document)

Rusch-Feja , Dianna. *The Open Archives Initiative and the OAI Protocol for Metadata Harvesting: rapidly forming a new tier in the scholarly communication infrastructure***Learned Publishing**. Vol.15, No.3, 2002, p. 179-186.

Shearer, K.: *The Open Archives Initiative : Developing an Interoperability Framework for Scholarly Publishing* ; ***CARL/ABRC Backgrounder Series*** , *No.5, March 2002.*

Tennant, Roy. *Metadata As If Libraries Depended on It*. **Library Journal**, 15 April 2002.

Van de Sompel, Herbert. *The OAI Protocol for Metadata Harvesting: an Update. Presentation at the CNI Fall 2002 Task Force Meeting*, Washington DC.

Van de Sompel, Herbert. *Systems for Scholarly Communication & the OAI-PMH. Lecture at the Frye Leadership Institute, Emory University*, Atlanta GA. June 11th 2002.

Van de Sompel, Herbert, Michael Nelson, Simeon Warner *The Open Archives Initiative Protocol for Metadata Harvesting*, v.2.0(beta). Editors: Carl Lagoze,. 2002.

Electronic journals:

Brown, Genevieve and Beverly Irby. *Fourteen Lessons: Initiating and Editing an Online Professional Refereed Journal* . **The Journal of Electronic Publishing** , No.8, August 2002.

Car, Leslie. *The Pressures Towards Open Access*. Presentation given at September 13 ALPSP conference, Open Access Journals --Will They Fly?.(Powerpoint document)

Carim, Lara. *Serial killers: how great is the e-print threat to periodicals publishers?* **Learned Publishing**. Vol.15, No.2, 2002, p.153 –155

Crawford, Walt. *Free electronic refereed journals: getting past the arc of enthusiasm*. **Learned Publishing**. Vol.15, No.1, 2002,p.117-123

Crawford, Walt. *Electronic Access to Scientific Articles: Another Perspective*. **Econtent,** May 2002.

Crow, Raym. *Converting an Existing Journal to Open Access* Presentation given at September 13 ALPSP conference, Open Access Journals --Will They Fly?.(Powerpoint document)

Foster, Andrea. *Second Thoughts on 'Bundled' E-Journals*. **Chronicle of higher education**, 20th September, 2002.

Ginn, Claire. *Calculating pricing models choices: rising to the challenge* **Learned Publishing**. Vol.15, No.3, 2002, p.199-203.

Hitchcock , Steve The Emerging infrastructure of Scholarly Communication Presented at We can't go on like this: the future of journals, *ALPSP International Learned Journals Seminar*, London. April 2002. (Powerpoint presentation)

Kircz Joost G. *New practices for electronic publishing 2: New forms of the scientific paper* **Learned Publishing**. Vol.15, No.1, 2002,p.27-32.

Kling, Rob;Geoff McKim and Lisa Spector. *Locally Controlled Scholarly Publishing Via the Internet: The Guild Model* **The Journal of Electronic Publishing**:Vol. 8, No.1; August 2002.

Kling, Rob Lisa Spector, and Geoff McKim. *Locally Controlled Scholarly Publishing via the Internet: The Guild Model*. **The Journal of Electronic Publishing**, No.8, August 2002.

McElroy, Emily. *Dos and don'ts for electronic journal management: some advice to publishers* **Learned Publishing**. Vol.15, No.1, 2002,p 125 –128

Morgan, Cliff. *Free electronic refereed journals Learned Publishing* **Learned Publishing**. Vol.15, No.3, 2002, p 238-239.

Powell, Andrea. *Linking to full text: the secondary publisher's perspective* **Learned Publishing,** Vol 15, No 4, October 2002, p 267-ff

Quint, Barbara. *BioMed Central Strengthens Research Library Connections*. **Information Today** , 20 May 2002.

Report from the UK Office of Fair Trading. *The Market for Scientific, Technical and Medical Journals*

Richardson, Martin. *Alternatives to Open Access* Presentation given at September 13 ALPSP conference, Open Access Journals --Will They Fly?.(Powerpoint document)

Sathe Nila A. ; Jenifer L. Grady and Nunzia B. Giuse. *Print versus electronic journals: a preliminary investigation into the effect of journal format on research processes*. **J Med Libr Assoc. ,** Vol 90, No.2, April 2002, p. 235–243.

Suber, Peter. *Open Access to the Scientific Journal Literature.* **Journal of Biology**, Vol.1, No.1, June 2002. pp3 ff.

Usher, Nicole B. *Scientists Try Free Journal Access*. **Harvard Crimson**, 1 March 2002

Velterop, Jan. *The Commercial Case for Open Access* Presentation given at September 13 ALPSP conference, Open Access Journals --Will They Fly?.(Powerpoint document)

Worlock ,Kate. *Electronic journals: user realities - the truth about content usage among the STM community* **Learned Publishing**. Vol.15, No.3, 2002, p.223 –226.

Archives, repositories and digital libraries.

Anderson Kent *The useful archive* **Learned Publishing**. Vol.15, No.1, 2002,p.85-89.

Breaks ,Michael. *Building the hybrid library: a review of UK activities* **Learned Publishing**. Vol.15, No.1, 2002,p.99-107.

Bromage; Sarah ; David Finkelstein and Alistair McCleery *Scottish Archive of Print and Publishing History Records* **Learned Publishing**. Vol.15, No.3, 2002, p.193-197

Chan, Leslie and Barbara Kirsop. *Open Archiving Opportunities for Developing Countries: towards equitable distribution of global knowledge* **Ariadne,** No.30, January 2002.

Cole, Timothy W.; Kaczmarek, Joanne; Marty, Paul F.; Prom, Christopher J.; Sandore, Beth and Sarah Shreeves. *The Illinois Open Archives Initiative Metadata*

Harvesting Experience **Museum and the Web :**The International Conference about Museums and the Web, 2002.

De Castro Neto Miguel and Ana Maria Ramalho Correia. *The Role of Eprint Archives in the Access to, and Dissemination of, Scientific Grey Literature: LIZA—A Case Study by the National Library of Portugal*

Gardner, Mike John MacColl and Stephen, Pinfield. *Setting up an institutional e-print archive* . **Ariadne**, Vol.11, No.31, April-2002.

Gutteridge, Christopher and Harnad, Stevan . *Applications, Potential Problems and a Suggested Policy for Institutional E-Print Archives .* August, 2002.

Marlino , Mary ;Michael Wright and Tamara Sumner. *Meta-Design of a Community Digital Library* **D-Lib** Vol. 8, No. 5, May 2002.

Mayfield ,Kendra . *College Archives 'Dig' Deeper* **Wired News,** Aug. 3, 2002.

Nixon , William. *The evolution of an institutional e-prints archive at the University of Glasgow* **Ariadne,** No.32. June July 2002.

Report on the Digital Preservation Planning Project. *The Yale Electronic Archive: One Year of Progress.*

Tenant, Roy. *Institutional Repositories*. **Library Journal**, 15 September 2002.

Different users, different needs.

Berry, R. Stephen, and Anne Simon Moffat, ed. *The Transition from Paper: Where Are We Going and How Will We Get There?* Cambridge, MA: American Academy of Arts & Sciences, Published online, 2001

Bollen , Johan Rick Luce *Evaluation of Digital Library Impact and User Communities by Analysis of Usage Patterns* **D-Lib** , Vol.8, No.6, June 2002.

Brusilovsky, Peter; Paul DeBra; Ricardo Conejo; eds. *Adaptative Hypermedia and Adaptative Web-Bases Systems. Proceedings of the Second International Conference, AH 2002, Malaga, Spain, May 2002.* Berlin: Springer, 2002, 615p.

Case , Mary. *Building the Scholars' Commons* . ALA Annual Conference,16 June 2002.

Young, Jeffrey R. *'Superarchives' Could Hold All Scholarly Output*. **The Chronicle of Higher Education**, 5 July 2002:

Fytton, Rowland *What do users want?* **Learned Publishing**. Vol.15, No.1, 2002,p 83-84

Gardner, Mike; John MacColl and **Stephen Pinfield** *Setting up an institutional e-print archive* **Ariadne,** No.31 April 2002

Genova, Francoise. *Online Information in Astronomy* Presentation made at the OAF Workshop in Pisa, May 2002. (Powerpoint document)

Goldschmidt-Clermont , Luisella. *Communication Patterns in High-Energy Physics* **CERN Webzine**, No.6, March 2002.

Harrison, Teresa M. and Timothy Stephen. *Intensive Disciplinarity in Electronic Services for Research and Education:Building Systems Responsive to Intellectual Tradition and Scholarly Culture* . **The Journal of Electronic Publishing** , No.8, August 2002.

Holtkamp, I. , Q. Hong, F. Knudson, X. Liu, K. Maly, M. Nelson and M. Zubair. (May 2002) *Federated Searching Interface Techniques for Heterogeneous OAI Repositories*. **Journal of Digital Information, Vol.**2, No. 4, May 2002.

Johnson ,Rick. *Working together to reform scholarly communication* **College & Research Libraries News ,** No. 63, October 2002

Lischka ,Konrad.*Der Geist, der aus der Flasche kam* . **Telepolis** , March 16 2002.

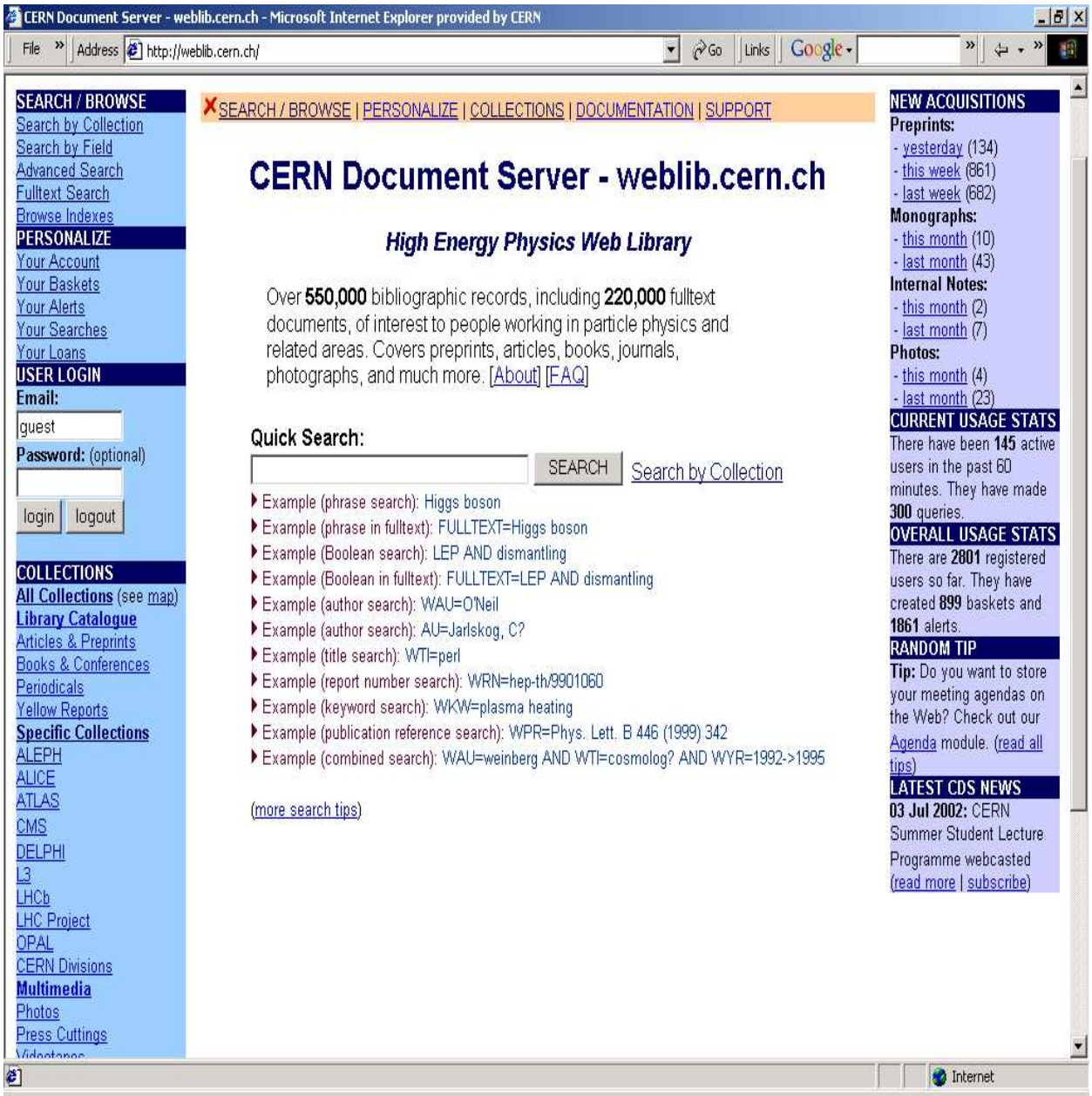
Tenopir, Carol and Donald W. King. *Reading behaviour and electronic journals*. **Learned Publishing,** Vol 15, No 4, October 2002, p 259-ff

The Knight Higher Education Collaborative . *Op. cit.: publishing in the humanities and social sciences* **Learned Publishing**. Vol.15, No.3, 2002, p.205- 216

Annexes

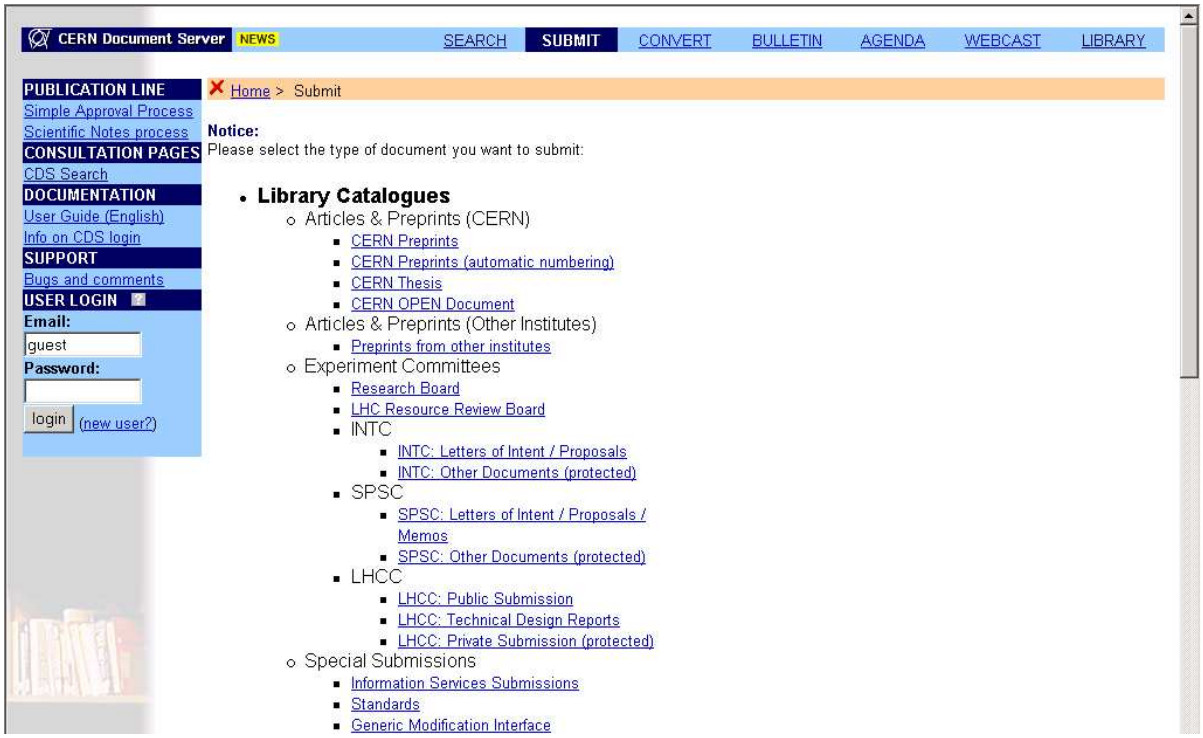
Annexe 1 CDS

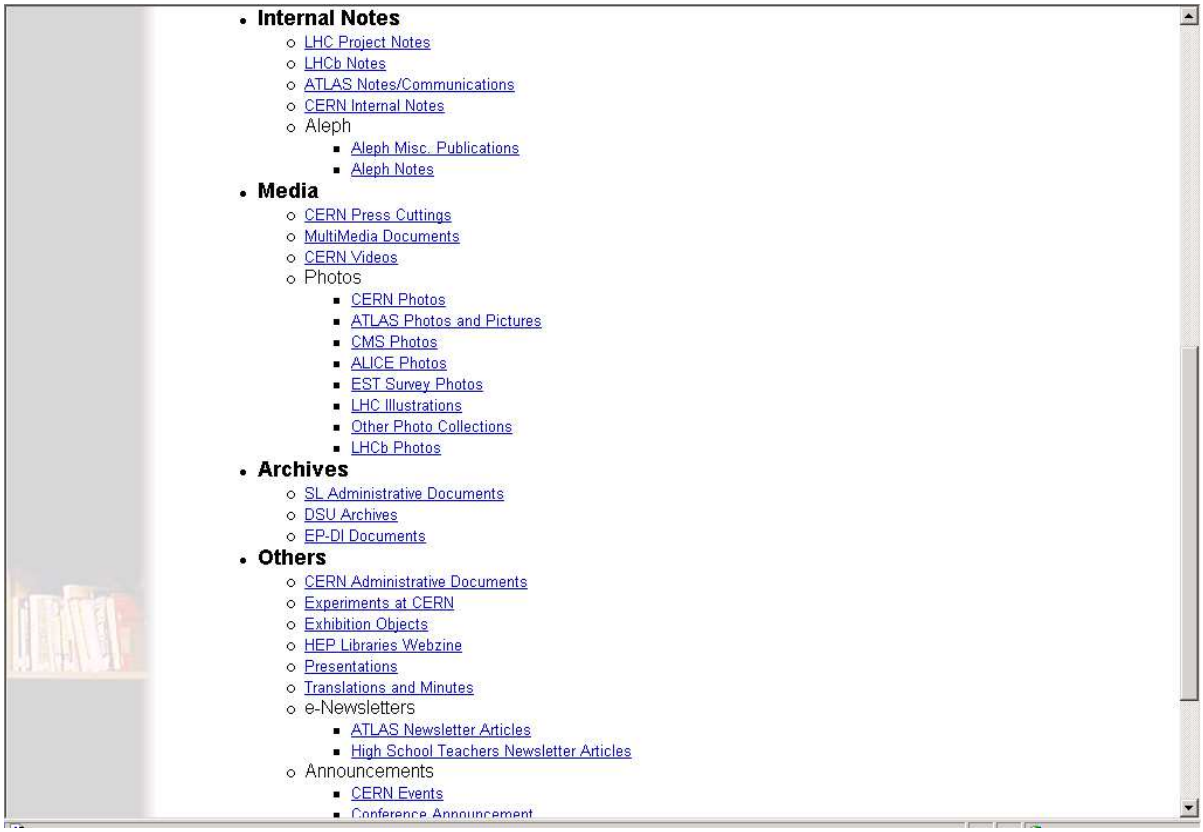
Annexe 1.1 Interface d'accueil de CDS



Annexe 1.2 Nouvelle interface de CDS

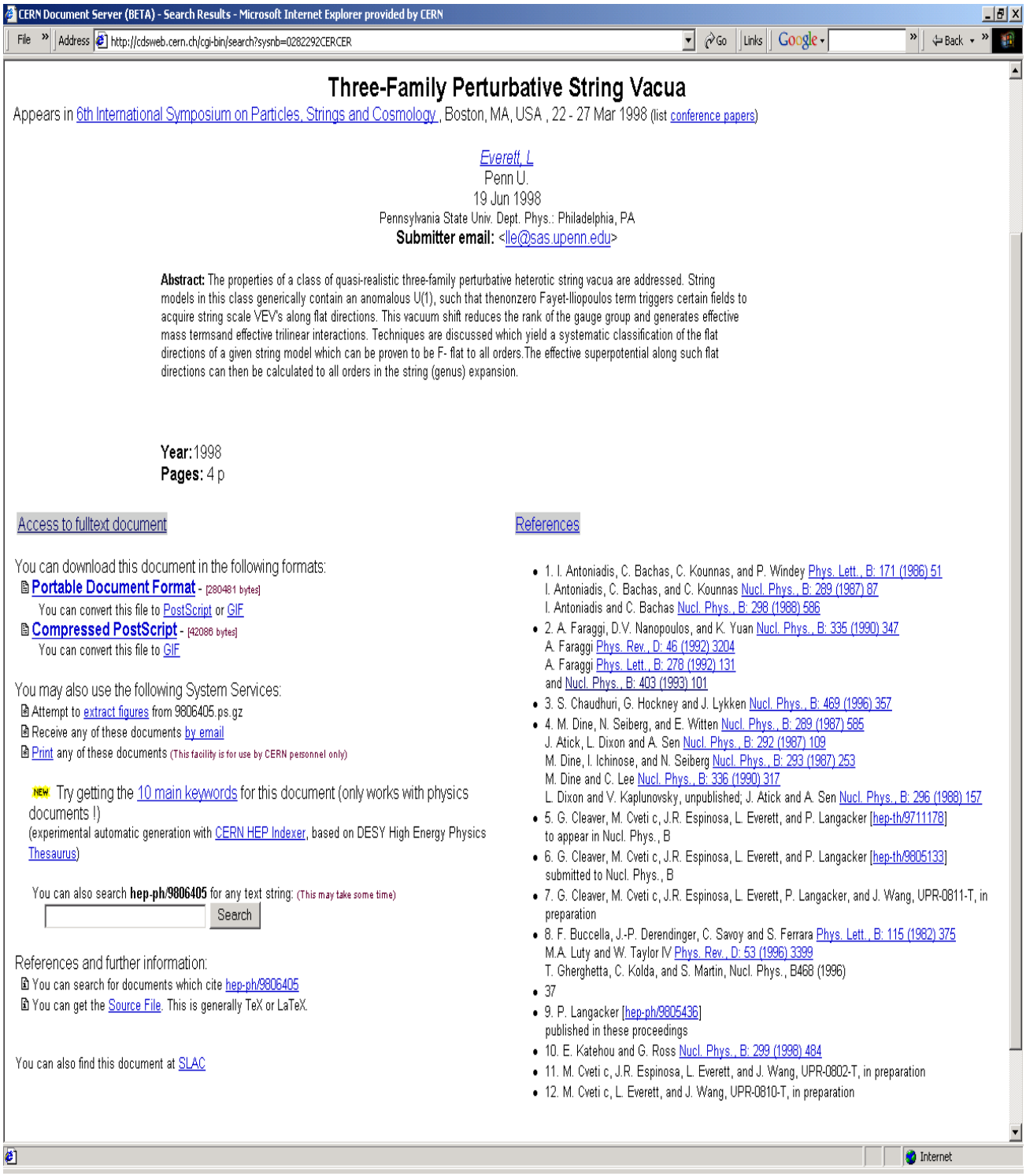
Annexe 1-3 Interface de soumission de CDS





Annexe 2 Documents produits

Annexe 2-1 Notice de CDS



CERN Document Server (BETA) - Search Results - Microsoft Internet Explorer provided by CERN

File » Address http://cdsweb.cern.ch/cgi-bin/search?sysnb=0282292CERCER Go Links Google Back »

Three-Family Perturbative String Vacua

Appears in 6th International Symposium on Particles, Strings and Cosmology , Boston, MA, USA , 22 - 27 Mar 1998 (list conference papers)

Everett, L
Penn U.

19 Jun 1998

Pennsylvania State Univ. Dept. Phys.: Philadelphia, PA
Submitter email: <lle@sas.upenn.edu>

Abstract: The properties of a class of quasi-realistic three-family perturbative heterotic string vacua are addressed. String models in this class generically contain an anomalous U(1), such that thenonzero Fayet-Iliopoulos term triggers certain fields to acquire string scale VEV's along flat directions. This vacuum shift reduces the rank of the gauge group and generates effective mass termsand effective trilinear interactions. Techniques are discussed which yield a systematic classification of the flat directions of a given string model which can be proven to be F- flat to all orders.The effective superpotential along such flat directions can then be calculated to all orders in the string (genus) expansion.

Year: 1998
Pages: 4 p

Access to fulltext document

You can download this document in the following formats:

Portable Document Format - [280481 bytes]
You can convert this file to PostScript or GIF

Compressed PostScript - [42086 bytes]
You can convert this file to GIF

You may also use the following System Services:

Attempt to extract figures from 9806405.ps.gz
Receive any of these documents by email
Print any of these documents (This facility is for use by CERN personnel only)

NEW Try getting the 10 main keywords for this document (only works with physics documents !)
(experimental automatic generation with CERN HEP Indexer, based on DESY High Energy Physics Thesaurus)

You can also search **hep-ph/9806405** for any text string: (This may take some time)
Search

References and further information:

You can search for documents which cite hep-ph/9806405
You can get the Source File. This is generally TeX or LaTeX.

You can also find this document at SLAC

References

- 1. I. Antoniadis, C. Bachas, C. Kounnas, and P. Windey Phys. Lett., B: 171 (1986) 51
  I. Antoniadis, C. Bachas, and C. Kounnas Nucl. Phys., B: 289 (1987) 87
  I. Antoniadis and C. Bachas Nucl. Phys., B: 298 (1988) 586
- 2. A. Faraggi, D.V. Nanopoulos, and K. Yuan Nucl. Phys., B: 335 (1990) 347
  A. Faraggi Phys. Rev., D: 46 (1992) 3204
  A. Faraggi Phys. Lett., B: 278 (1992) 131
  and Nucl. Phys., B: 403 (1993) 101
- 3. S. Chaudhuri, G. Hockney and J. Lykken Nucl. Phys., B: 469 (1996) 357
- 4. M. Dine, N. Seiberg, and E. Witten Nucl. Phys., B: 289 (1987) 585
  J. Atick, L. Dixon and A. Sen Nucl. Phys., B: 292 (1987) 109
  M. Dine, I. Ichinose, and N. Seiberg Nucl. Phys., B: 293 (1987) 253
  M. Dine and C. Lee Nucl. Phys., B: 336 (1990) 317
  L. Dixon and V. Kaplunovsky, unpublished; J. Atick and A. Sen Nucl. Phys., B: 296 (1988) 157
- 5. G. Cleaver, M. Cveti c, J.R. Espinosa, L. Everett, and P. Langacker [hep-th/9711178]
  to appear in Nucl. Phys., B
- 6. G. Cleaver, M. Cveti c, J.R. Espinosa, L. Everett, and P. Langacker [hep-th/9805133]
  submitted to Nucl. Phys., B
- 7. G. Cleaver, M. Cveti c, J.R. Espinosa, L. Everett, P. Langacker, and J. Wang, UPR-0811-T, in preparation
- 8. F. Buccella, J.-P. Derendinger, C. Savoy and S. Ferrara Phys. Lett., B: 115 (1982) 375
  M.A. Luty and W. Taylor IV Phys. Rev., D: 53 (1996) 3399
  T. Gherghetta, C. Kolda, and S. Martin, Nucl. Phys., B468 (1996)
- 37
- 9. P. Langacker [hep-ph/9805436]
  published in these proceedings
- 10. E. Katehou and G. Ross Nucl. Phys., B: 299 (1998) 484
- 11. M. Cveti c, J.R. Espinosa, L. Everett, and J. Wang, UPR-0802-T, in preparation
- 12. M. Cveti c, L. Everett, and J. Wang, UPR-0810-T, in preparation

Internet

Annexe 2.2 Application des propositions à une notice

Three-Family Perturbative String Vacua

Appears in 6th International Symposium on Particles, Strings and Cosmology , Boston, MA, USA , 22 - 27 Mar 1998 (list conference papers)

Everett, L Penn U. 19 Jun 1998 Pennsylvania State Univ. Dept. Phys.: Philadelphia, PA

Submitter email: <lle@sas.upenn.edu>

Abstract: The properties of a class of quasi-realistic three-family perturbative heterotic string vacua are …

Year:1998

Pages: 4 p

Type de document: Pré publication, rapport, auto archivage, e-print

P Lien vers la version Publiée

Certification: ★★★ selon l'échelle établie

Access to fulltext document

You can **download** this document in the following formats:

Portable Document Format - [280481 bytes]

You can convert this file to PostScript or GIF

Compressed PostScript - [42086 bytes]

- You can convert this file to GIF
- You can get the Source File. This is generally TeX or LaTeX.
- You can also find this document at SLAC

Tools

D Number of downloads of the document

C Number of citations and document citing this document

10 main keywords for this document (only works with physics documents !) (experimental automatic generation with CERN HEP Indexer, based on DESY High Energy Physics Thesaurus)

References

- 1. I. Antoniadis, C. Bachas, C. Kounnas, and P. Windey Phys. Lett., B: 171 (1986) 51 I. Antoniadis, C. Bachas, and C. Kounnas Nucl. Phys., B: 289 (1987) 87 I. Antoniadis and C. Bachas Nucl. Phys., B: 298 (1988) 586
- 2. A. Faraggi, D.V. Nanopoulos, and K. Yuan Nucl. Phys., B: 335 (1990) 347A. Faraggi Phys. Rev., D: 46 (1992) 3204 A. Faraggi Phys. Lett., B: 278 (1992) 131 and Nucl. Phys., B: 403 (1993) 101
- 3. S. Chaudhuri, G. Hockney and J. Lykken Nucl. Phys., B: 469 (1996) 357
- 4. M. Dine, N. Seiberg, and E. Witten Nucl. Phys., B: 289 (1987) 585 J. Atick, L. Dixon and A. Sen Nucl. Phys., B: 292 (1987) 109 M. Dine, I. Ichinose, and N. Seiberg Nucl. Phys., B: 293 (1987) 253 M. Dine and C. Lee Nucl. Phys., B: 336 (1990) 317 L. Dixon and V. Kaplunovsky, unpublished; J. Atick and A. Sen Nucl. Phys., B: 296 (1988)

High Energy Physics Thesaurus)

Full TEXT search

You can also search **hep-ph/9806405** for any text string: (This may take some time)

Search

System Services:

Attempt to extract figures from 9806405.ps.gz

Receive any of these documents by email

Print any of these documents (This facility is for use by CERN personnel only

Format: HTML | HTML MARC | XML DC | XML MARC

157

- 5. G. Cleaver, M. Cveti c, J.R. Espinosa, L. Everett, and P. Langacker [hep-th/9711178] to appear in Nucl. Phys., B
- 6. G. Cleaver, M. Cveti c, J.R. Espinosa, L. Everett, and P. Langacker [hep-th/9805133] submitted to Nucl. Phys., B
- 7. G. Cleaver, M. Cveti c, J.R. Espinosa, L. Everett, P. Langacker, and J. Wang, UPR-0811-T, in preparation
- 8. F. Buccella, J.-P. Derendinger, C. Savoy and S. Ferrara Phys. Lett., B: 115 (1982) 375
  M.A. Luty and W. Taylor IV Phys. Rev., D: 53 (1996) 3399
  T. Gherghetta, C. Kolda, and S. Martin, Nucl. Phys., B468 (1996)
- 9. P. Langacker [hep-ph/9805436] published in these proceedings
- 10. E. Katehou and G. Ross Nucl. Phys., B: 299 (1998) 484
- 11. M. Cveti c, J.R. Espinosa, L. Everett, and J. Wang, UPR-0802-T, in preparation
- 12. M. Cveti c, L. Everett, and J. Wang, UPR-0810-T, in preparation

Annexe 2-3 Enquête portant sur la certification des documents

Bonjour,

Je suis actuellement stagiaire au sein de la bibliothèque du CERN et je travaille sur de nouvelles options de classement des documents en ligne. Dans le cadre de l'amélioration de nos services aux usagers et a la communauté scientifique, nous travaillons sur une fonctionnalité ayant trait aux documents issus des principaux laboratoires de recherche en Physique.

Un certain nombre de documents provenant de votre institut sont présents dans notre base de données et nous aurions besoin de quelques renseignements supplémentaires afin de mieux les exploiter.

En effet, nous cherchons a inclure un nouveau champs de méta données qui ferait apparaître un niveau de "validation" des documents électroniques présents dans notre serveur de preprints et d'articles.

Pour ce faire, il nous serait grandement utile de connaître avec le plus de précision possible les procédures de validation des documents en vue de l'attribution d'un Report Number au moment de leur soumission et de leur mise en ligne en tant que preprint ou document auto-archive.

Nous avons préparé un certain nombre de questions, auxquelles nous vous serions extrêmement reconnaissants de consacrer quelques instants si cela vous est possible.

Les voici :

Le document est-il vérifié?

Le document est-il vérifié par des membres du département, de la division, ou de l'unité?

Le document est-il vérifié par des membres extérieurs?

Le document est-il vérifié par un commission composée de membres du département, de la division, de l'unité et de membres extérieurs?

Le document est-il vérifié par un autre type de commission? Dans ce cas pourriez- vous le décrire?

Bien entendu, nous vous communiquerons les résultats de cette enquête dans les meilleurs délais.

D'avance, nous vous remercions de l'attention que vous saurez porter à ces quelques questions

Cordialement,

Dear Sir or Madam

I am currently an intern at the CERN library and I am working on improving the library's online service to the user and to the scientific community thus we are trying to include a new option in our ranking procedures.

The main idea is to rank electronic documents (pre-prints or self-archived documents not designed for publishing) according to the process of reviewing they went through before being online.
As a number of documents in our database come from your institute, we would like you to provide us with some more information, if possible. Our main interest lies in the different reviewing procedures used in your institute.

We would greatly appreciate if you could answer a few questions:

Does the document go through any process of validation?

Does the committee which validate the document include the head of the department/division/unit and some selected scholars from the institute?

Does the committee which validate the document include members of the department/division/unit and some selected scholars from other institutes?

Does the committee which validate the document include external scholars and specialists?

Do you use any other kind of validation process? If you do, could you describe it?

Of course, we would provide you with the results as soon as possible and would be at your disposal for any further question.

Your collaboration will be of great use, and we hope you will find this idea interesting enough to spend some time on it.

Best regards

Annexe 2.4 Les chercheurs et la recherche documentaire

Ce texte reprend les grandes lignes d'un article en préparation sur le comportement des chercheurs dans le cadre de la recherche documentaire.

Il existe beaucoup de travaux portant sur les pratiques documentaires des chercheurs ; cependant peu s'intéressent à l'utilisation faite par les chercheurs des ressources en ligne. Pourtant il serait nécessaire de s'interroger sur leurs méthodes et leurs besoins réels. Mes entretiens avec Michelangelo Mangano, physicien théoricien, et Franck Laloe, physicien et responsable d'un serveur dépendant du CNRS, HAL , m'ont permis d'appréhender une réalité dont j'avais été coupée . Les chercheurs ne sont pas des bibliothécaires et ne souhaitent pas le devenir., les interfaces trop lourdes, les recherches multi critères nécessitant des équations complexes les font en général fuir. Leur approche est pragmatique : ils ont besoin de trouver rapidement les documents nécessaires et de la manière la plus intuitive et la plus simple possible. En général et dans le domaine de la physique des hautes énergies en particulier, les scientifiques ne changent pas de sujet de recherche tous les ans ; ils choisissent une spécialisation et y restent pendant des années. De fait, ils appartiennent à une communauté d'environ 200 à 300 spécialistes du sujet qu'ils traitent ; communauté avec ses affinités, ses dissensions, ses inimitiés qui se font et se défont au fil des colloques, des collaborations et des publications. Il est donc extrêmement rare de voir un chercheur, sauf un étudiant débutant sa thèse ou entrant dans un domaine, aborder une recherche sans une idée de ce qu'il cherche : il connaîtra soit l'auteur rencontré ou entendu dans une conférence, soit le titre d'un article mentionné par un collègue, soit le laboratoire qui effectue les expériences dont il a besoin. Cela est corroboré par l'analyse des requêtes stockées sur CDS qui montre que 95% des requêtes sont simples et ne portent que sur les champs les plus communs (auteur, date, report number, mot du titre).De fait, les requêtes stockées faisant plus d'une ligne et utilisant de nombreux opérateurs sont souvent, mais pas exclusivement, celles des informaticiens testant le système…

D'autre part, les physiciens sont abonnés à des listes de diffusions quotidiennes qui leur soumettent les nouveaux articles dans leur domaine : les titres d'environ 100

articles hebdomadaires sont parcourus rapidement et deux ou trois sont imprimés au maximum. Les physiciens font donc une veille documentaire quotidienne portant sur une ou deux sources au maximum et font rarement appel aux moteurs de recherche des bibliothèques sauf pour y localiser des documents précis. Que l'article soit publié et donc reconnu par un comité de pairs n'influe pas sur l'intérêt que le chercheur lui portera : le prestige d'une publication n'intéresse que l'auteur et les universités qui se basent dessus pour recruter. Cet aspect est donc de bon augure pour l'auto archivage, puisque le contenu est privilégié :les pré publications ne sont donc pas perçues, dans le domaine de la physique, comme de moins bons articles mais comme des instruments de recherche indispensables et rapidement disponibles.

D'autre part, il est à noter que nombre de physiciens ne se fient ,en général, qu'à leur appréciation personnelle d'un document :peu leur importe qu'un comite d'inconnus ou de spécialistes ait validé l'article. En revanche, un article recommandé par un ami ou un collègue aura toutes les chances d'attirer leur attention. C'est là un élément clef de la recherche scientifique et qui rejoint la notion de communauté : un moteur de recherche quelque puissant qu'il soit ne pourra jamais détecter les inimitiés ou les préférences de l'utilisateur : un article pourra remplir tous les critères de qualité et de pertinences, il aura peu de chances d'être lu s'il émane d'un institut ou d'un chercheur que l'utilisateur n'apprécie pas…Cela peut paraître trivial mais correspond à une réalité que j'ai eu maintes fois l'occasion de rencontrer.

Enfin, le poids des habitudes est une donnée essentielle de la communauté scientifique que j'ai observée : depuis 11 ans, les physiciens de la division théorie du CERN soumettent leurs preprints au serveur ArXiv et non au serveur du CERN. Le CERN est donc obligé de récupérer tous les jours les données de ses propres chercheurs sur un site extérieur. Le serveur du CERN offre un traitement des méta données et une interface de recherche nettement plus élaborés que ceux fournis par Spire et ArXiv, cependant, les chercheurs sont habitués à cette interface, qui a eu l'immense avantage de ne pratiquement pas évoluer en 11 ans, ce qui lui donne évidemment toute sa valeur..

Il semble que l'aspect comportemental pourrait fournir bon nombre de réponses aux questions que l'ont se posent sur l'auto archivage . Il apparaît nécessaire de faire une étude portant sur les chercheurs dans différentes disciplines afin de mettre en

adéquation les besoins et les pratiques avec les développements informatiques et bibliothéconomiques.