

L'archive institutionnelle de l'Ecole Polytechnique
Fédérale de Lausanne : état actuel et perspectives

Cet article fait suite à la présentation d'Infoscience au cours de la Semaine du Document
Numérique 2006, organisée à Fribourg, Suisse, du 15 au 19 septembre 2006

David Aymonin

Service de l'Information scientifique et des bibliothèques, EPFL -
Ecole Polytechnique Fédérale de Lausanne, VPAA-DAF-SISB, 1015
Lausanne, Suisse David.Aymonin@epfl.ch

Pierre Crevoisier

Knowledge and Information Service, EPFL - Ecole Polytechnique
Fédérale de Lausanne, VPPL-KIS, 1015 Lausanne Suisse
Pierre.Crevoisier@epfl.ch

Frédéric Gobry

Google Ireland, Barrow Street, Dublin 4, Ireland gobry@google.com

archive institutionnelle, Open access, facteur d'impact des
publications, soutien à la recherche, gestion, diffusion et
réappropriation de l'information scientifique et technique,
édition électronique, archive électronique, gestion projet,
analyse conceptuelle, description système, interface utilisateur,
Suisse, EPFL (Ecole polytechnique fédérale de Lausanne)

electronic publishing, digital archive, institutional repository,
project management, conceptual analysis, system description, user
interface, Suisse, EPFL (Ecole polytechnique fédérale de Lausanne)

L'Ecole polytechnique fédérale de Lausanne s'est dotée en 2004
d'une archive institutionnelle permettant la collecte et la
valorisation de son patrimoine scientifique et humain. Cet article
décrit les objectifs et les réalisations de ce projet en insistant
sur les principes de conception du service proposé aux chercheurs
et à l'institution, et à leur adaptation en fonction des
enseignements tirés de l'expérience. Les facteurs influençant
l'adoption, par les utilisateurs, des outils et services proposés
sont analysés.

Introduction

L'EPFL est une université technologique qui accueille aujourd'hui près de 6.000 étudiants, dont plus de 1.000 doctorants. 3.000 chercheurs et scientifiques y développent des enseignements et des recherches en mathématiques, physique, chimie, management de la technologie, sciences de l'ingénieur, sciences de la vie, architecture.

Dans le nouvel environnement universitaire de la réforme de Bologne¹ et de la globalisation de l'enseignement supérieur, et s'appuyant sur sa bonne réputation scientifique, l'Ecole s'est fixé l'objectif de se situer parmi les meilleures en Europe et dans le monde, pour la qualité de sa recherche comme de son enseignement.

En 2004, pour faire suite à une demande de la Direction de l'Ecole de remplacer et renouveler le rapport scientifique produit entre 2002 et 2004, le service informatique et la bibliothèque ont reçu le mandat de mettre en place et tester une solution destinée à collecter et valoriser le patrimoine scientifique de l'EPFL.

Après 6 mois consacrés à la comparaison de solutions existantes, de test et d'installation d'un outil logiciel adapté, Infoscience, à la fois service et serveur de l'archive institutionnelle, était ouvert.

¹ La réforme dite « de Bologne » prévoit au sein de l'Union européenne des niveaux d'étude standards dans l'enseignement supérieur (LMD en France, Bachelor-Master-Doctorat en Suisse) et un système de crédits d'enseignement transférables (ECTS), reconnus mutuellement entre les pays.

Le logiciel open source CDS Invenio [1] (anciennement CDSware, <http://cdsware.cern.ch/invenio/index.html>), développé par le Centre Européen pour la Recherche Nucléaire (CERN), a été choisi pour gérer la plateforme technique servant de support à ce service. En septembre 2006 Infoscience présente le profil de plus de 1.200 chercheurs et les publications de 129 laboratoires, sur les 240 qui composent l'EPFL. 26 000 publications y sont signalées, 10.000 Fulltexts y sont stockés, dont l'ensemble des thèses de l'EPFL. De plus Infoscience est destiné à devenir l'interface d'interrogation du catalogue collectif des bibliothèques de l'EPFL. C'est donc une plateforme de services liés au patrimoine scientifique humain (constitué de profils des personnes), produit (constitué des publications scientifiques et pédagogiques) ou acquis (constitué des documents reçus ou achetés) que vise à offrir Infoscience. L'ensemble de ces contenus, services et fonctions est décrit plus en détail dans la suite de ce document. En matière d'archives institutionnelles, chaque cas est unique, selon l'environnement technique, humain, financier et administratif de chaque université. Il ne serait donc pas très utile d'insister sur les aspects historiques de la mise en œuvre d'Infoscience ni sur la machinerie interne. Par contre, il nous a semblé utile de présenter ici les fonctionnalités d'Infoscience, telles que nous avons pensé les implanter, les principes à la base du projet qui ont guidé le dialogue avec les différents partenaires et « actionnaires », et les réflexions et enseignements tirés de l'expérience (parfois déroutante) de sa mise en place. Nous souhaitons ainsi proposer un apport méthodologique à la communauté des spécialistes de l'information, bibliothécaires, informaticiens, chercheurs, ayant à implanter et gérer une archive institutionnelle.

Les « dimensions » de l'EPFL définissent un contexte particulier

L'EPFL se situe en dehors de la ville de Lausanne, à quelques centaines de mètres du lac Léman et jouxte l'Université de Lausanne (UNIL).

III. 1 : Vue d'ensemble de l'EPFL

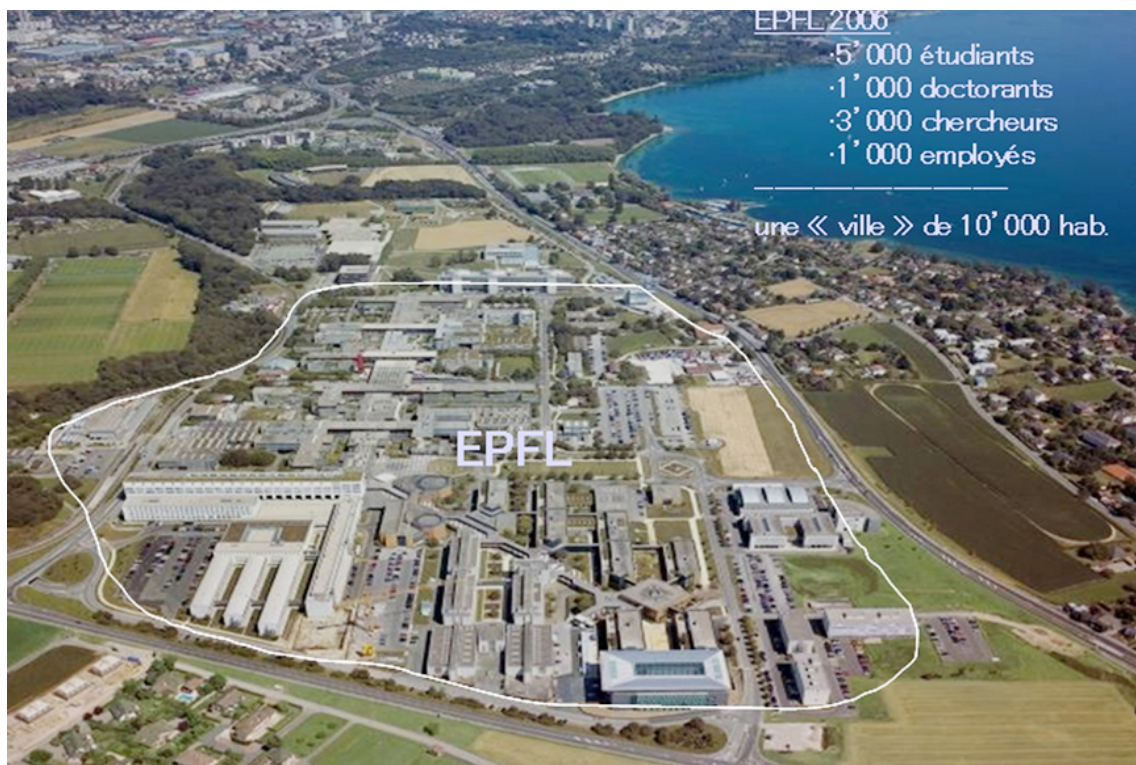


Figure 1 : Vue d'ensemble de l'EPFL

En 2006, l'EPFL compte une population d'environ 10.000 personnes dont 6.000 étudiants et doctorants, ainsi que 4.000 chercheurs et employés. Classée en août 2006 par Newsweek 26^e

meilleure université et 1^{re} université francophone au monde [2], elle vise un niveau de performance particulièrement élevé, en terme de qualité de sa recherche comme de son enseignement.

Les domaines d'enseignement et de recherche de l'EPFL sont répartis en 5 facultés et 2 collèges.

- La **Faculté des Sciences et techniques de l'Ingénieur (STI)** : Génie électrique et électronique, génie mécanique, matériaux et microtechnique ;
- La **Faculté Environnement naturel, architectural et construit (ENAC)** : Architecture, Génie civil, Sciences et ingénierie de l'environnement ;
- La **Faculté Informatique et Systèmes de communication (IC)** ;
- La **Faculté des Sciences de Base (SB)** : mathématiques, chimie et génie chimique et physique ;
- La **Faculté des sciences de la vie (SV)** ;
- Le **Collège du Management** : combinaison des disciplines traditionnelles liées aux Ecoles de Commerce, aux programmes de génie industriel, et aux facultés des sciences sociales ;
- Le **Collège des Humanités**, programme d'enseignement en sciences humaines et sociales (SHS) en partenariat avec les Universités de Lausanne et Genève.



Figure 2 : Facultés et sections, détail et implantation

Infoscience est à la fois un outil et une plateforme de services

Une visite guidée permet de se faire une idée concrète du contenu, des fonctionnalités de l'outil et des services proposés.

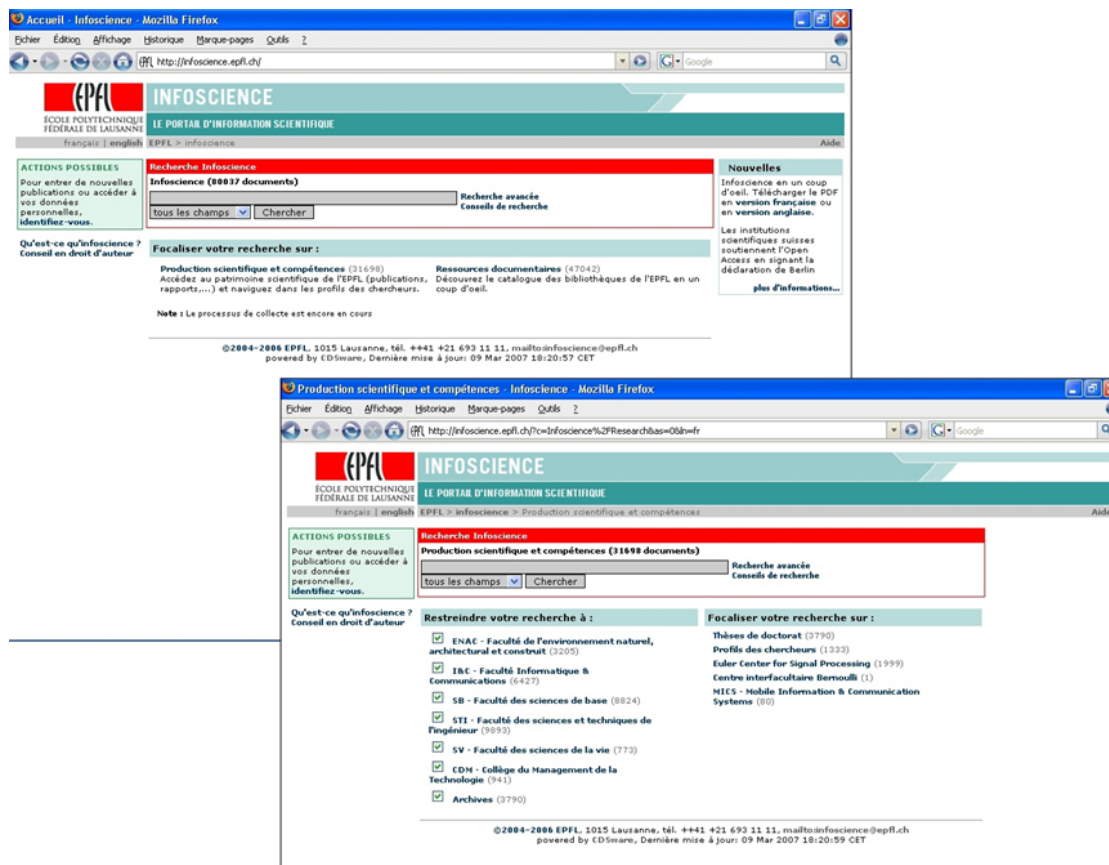


Figure 3 : Interfaces principales d'Infoscience.epfl.ch

Le contenu

Références et textes intégraux des publications scientifiques

Les types de documents acceptés et le statut des publications (pre- et post-print, peer reviewed, submitted, accepted) sont visibles dans l'écran de recherche avancée. Le fulltext des documents signalés peut être ajouté, et rendu accessible ou non au monde extérieur, selon le choix de l'auteur. Nous envisageons d'ouvrir au moissonnage OAI les publications dont le fulltext est disponible, tout en respectant le droit d'auteur.

Description et identification des personnes et de leurs compétences

Infoscience intègre les données des « Profils des personnes » saisis dans l'application [people@EPFL](http://people.epfl.ch) (<http://people.epfl.ch>). Cet annuaire évolué permet de présenter les curriculum vitae des personnels scientifiques et des étudiants doctorants. Il est ainsi possible d'explorer au travers de l'interface de recherche les compétences scientifiques de l'EPFL.

Collections de documents numérisés ou numériques

Les thèses de l'EPFL soutenues depuis 1920 ont toutes été numérisées, et sont accessibles via Infoscience [3]. Un processus automatique et obligatoire d'archivage des nouvelles thèses sous forme numérique a été mis en place pour toute l'Ecole depuis 2003. Les thèses dont les auteurs ont signé l'autorisation de mise en ligne sans restriction sont déclarées dans les moissonneurs OAI-PMH²

² Moissonnage OAI-MPH : récolte des métadonnées contenues dans une archive ou une base de données selon les règles de l'Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>

comme OAlster (<http://oaister.umdl.umich.edu/>) ou RERODOC (<http://doc.rero.ch/>). Nous envisageons de signaler les thèses dans The Networked Digital Library of Theses and Dissertations (NDLTD, <http://www.ndltd.org/index.en.html>), directement ou via le serveur d'archives de la Bibliothèque nationale suisse qui a la mission de collecter l'ensemble des thèses au niveau national (http://www.nb.admin.ch/slb/slb_professionnel/projektarbeit/00719/00723/00629/).

Certains enseignants ont commencé à signaler et charger leurs supports de cours, mais aucune politique de collecte systématique n'a encore été établie. À terme, nous prévoyons de développer la mise en ligne des travaux de master produits par les étudiants.

Le catalogue collectif des bibliothèques

Infoscience sert d'interface à l'interrogation des catalogues des plus importantes bibliothèques de l'EPFL. Actuellement les données des catalogues de 11 bibliothèques sont chargées. Nous attendons la mise en place d'un export OAI des données venant de 14 bibliothèques de l'EPFL membres du réseau de bibliothèques suisses NEBIS (<http://www.nebis.ch>). Le catalogue collectif sera alors complet.

Les principaux services

Ils s'appuient sur les fonctionnalités existantes du logiciel open source CDS Invenio (anciennement CDSware) développé par le CERN, complétées par des développements adaptés aux besoins de l'EPFL. Le CERN a accepté de partager avec l'EPFL le développement de CDS Invenio. Ce co-développement devrait bientôt être formalisé par un contrat de partenariat signé entre les deux institutions.

Les fonctions et services sont résumés dans une brochure disponible en anglais et en français sur le site <http://infoscience.epfl.ch>.

Interfaces de recherche Google-like et avancée

Par défaut, Infoscience propose une interface extrêmement simple permettant d'explorer

- les publications scientifiques et productions pédagogiques
- les profils des personnes (annuaire et biographies des chercheurs)
- les catalogues des bibliothèques

Pour des requêtes plus fines une interface de recherche avancée permet de combiner les critères habituels de recherche dans les bases bibliographiques, mais s'enrichit de critères de choix du statut de la publication (peer reviewed, pre ou post print, soumis ou accepté, ...). Les interfaces simple et avancée sont disponibles en anglais et en français, et s'appuient sur le moteur de recherche évolué interne à CDS Invenio (permettant la lemmatisation, le classement par pertinence, la recherche dans tous les champs – et à terme dans le texte intégral des documents).

Réutilisation des données par les chercheurs eux-mêmes, selon leurs besoins

Toutes les références bibliographiques peuvent être exportées dans divers formats (BibTeX, EndNote, HTML, XML) pour une intégration simplifiée dans des documents ou applications gérés par l'utilisateur : site web du laboratoire, formulaire de demande de financement, rapport ou document en cours de frappe, base bibliographique, curriculum vitae personnel, etc. Les requêtes peuvent être conservées et utilisées pour lancer des alertes ou gérer des collections virtuelles de documents.

III. 4 : intégration d'Infoscience dans un site web de laboratoire (LANOS)

The image shows two screenshots of a web browser displaying the LANOS website. The top screenshot shows the 'People at LANOS' page, which lists staff members categorized by their roles: PROFESSOR (Martin Hasler), SENIOR RESEARCH FELLOW (Bertrand Dutoit), POST DOCS (Joseph Duron, Heinz Köppl), PHD STUDENTS (Jugoslav Acimovic, Ali Ajdari Rad, Cristian Carmeli, Enno de Lange, Mahdi Jalili, Slobodan Kozic, Kumiko Oshima), ADMINISTRATIVE STAFF (Erika Raetz), and FORMER COLLABORATORS (Gianni Bagni, Norman Uwe Baier, Igor Belykh). Each entry includes a name, contact information, and a small icon.

The bottom screenshot shows the profile page for Bertrand Dutoit. It includes a photo, contact information (bertrand.dutoit@epfl.ch), and a list of 'MAIN PUBLICATIONS' with titles and links to full texts. The publications listed are:

- S. Stavrev, F. Grilli, B. Dutoit, and S. Ashworth. Comparison of the AC losses of BSCCO and YBCO conductors by means of numerical analysis. *Superconductor Science Technology*, 18:1300-1312, 2005. [Details] [Full Text]
- J. Duron, F. Grilli, L. Antognazza, M. Desroux, B. Dutoit, and O. Fischer. Finite-element modelling of YBCO fault current limiter with temperature dependent parameters. *Superconductor Science and Technology*.
- J. Duron, F. Grilli, B. Dutoit, and S. Stavrev. Modelling the E-J relation of high-Tc superconductors in an arbitrary current range. *Physica C*, 401(1-4):231-235, 2004. [Details]
- F. Grilli, S. Stavrev, B. Dutoit, S. Spreafico, R. Tebano, G. Coletta, F. Gomory, L. Frolek, and J. Touc. Numerical modelling of a HTS cable in AC regime. *Physica C*, 401(1-4):176-181, 2004. [Details]
- S. Stavrev, B. Dutoit, and N. Nibbio. Geometry considerations for use of Bi-2223(Ag tapes and wires with different models of Jc(B). *IEEE Transactions on Applied Superconductivity*, 12(3):1857-1865, 2002. [Details]
- F. Grilli, S. Stavrev, Y. Le Floch, M. Costa Bouzo, E. Vivot, I. Klutsh, G. Meunier, P. Tirador, and B. Dutoit. Finite Element Method Modelling of Superconductors: from 2D to 3D. *IEEE Transactions on Applied Superconductivity*, 15(issue 1):17-25, 2004. [Details]

The bottom screenshot also shows a search interface for publications, with fields for Author, Title, Year, Journal, Abstract, Type, Research Field, Research Subfield, Status, and Maximum Results. It includes a search button and a list of results.

Indexation des publications et profils de personne dans Google et Google Scholar pour une bonne visibilité sur le web.

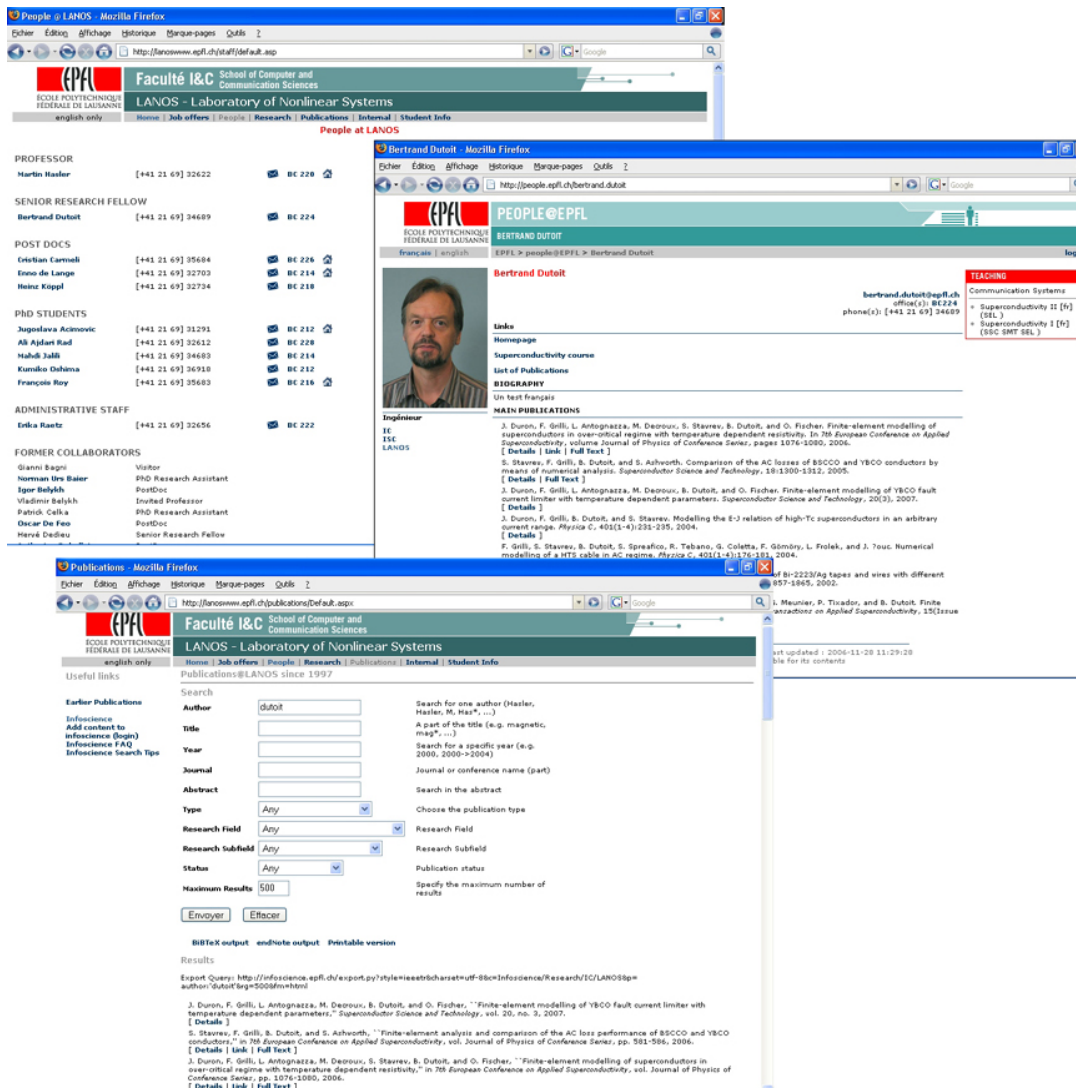


Figure 4 : Intégration d'Infoscience dans un site web de laboratoire (LANOS)

La structure des url et des fichiers a été optimisée pour donner de meilleurs résultats lors de l'indexation par le robot de Google. Ainsi les publications et les profils de personnes issus d'Infoscience apparaissent dans les premières réponses lorsque l'on interroge Google. D'autres outils et moteurs de recherche ont été contactés pour qu'ils indexent Infoscience : Scirus (<http://scirus.com>), Scopus (<http://www.scopus.com>), Web of Science, (<http://scientific.thomson.com/products/wos/>), Windows Academic live (academic.live.com/), OAlster, ... Les résultats avec Google et Google Scholar sont très bons, le dialogue est plus complexe avec d'autres acteurs commerciaux, et difficile avec Thomson Scientific (anciennement ISI), producteur de Web of Science.

Interface de gestion de la qualité

Pour faciliter la gestion de la qualité des données saisies par les différents utilisateurs du service (les principes d'alimentation sont décrits plus loin), un contrôle semi automatisé de la qualité des données a été mis en place (en version beta pour l'instant) afin de repérer les références et les documents posant problème : statut de publication incorrect, candidat doublon, variations de titre de journal, variation de graphie de noms d'auteurs, etc.

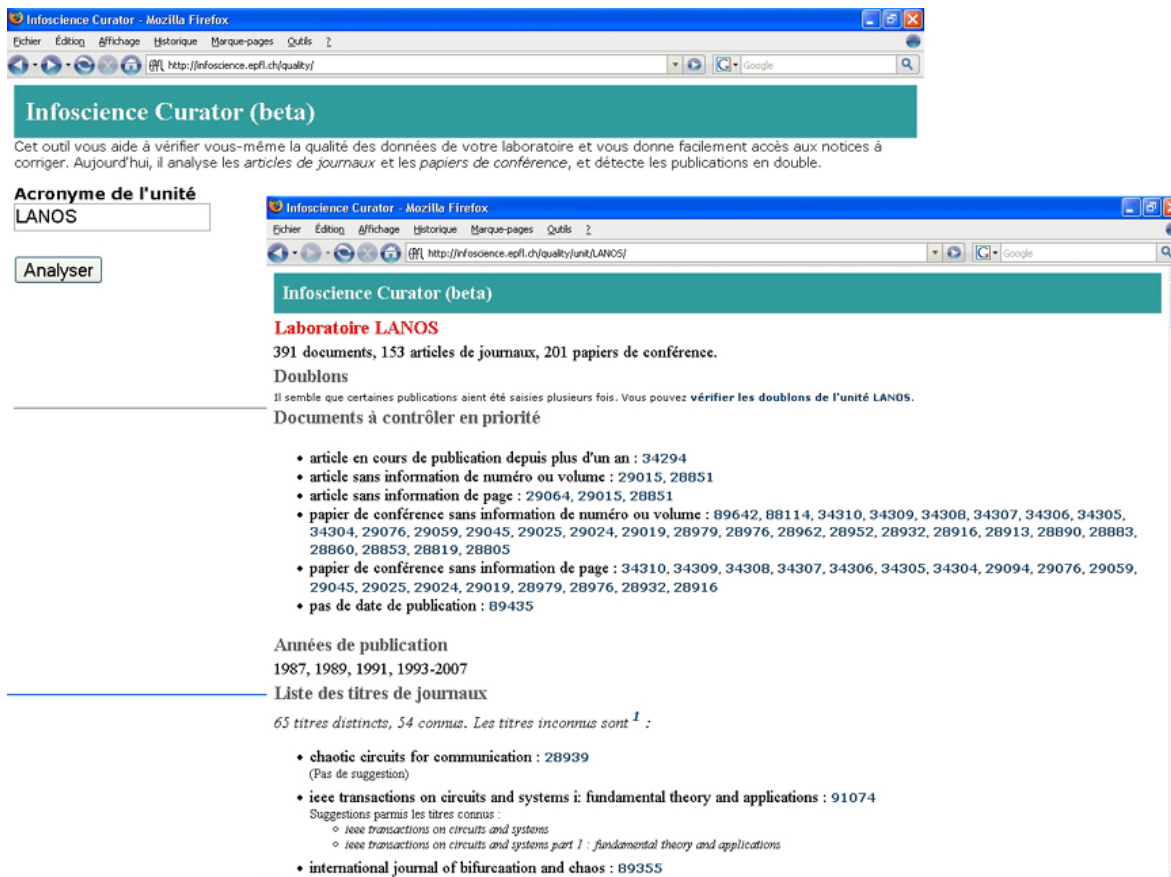


Figure 5 : Interface de gestion de la qualité (LANOS)

Conseil en droit d'auteur

Les auteurs pouvant soumettre eux-mêmes les références de leurs publications et charger les fichiers correspondants en pdf ou dans d'autres formats, il est vite apparu que des informations sur le droit d'auteur devaient apparaître au moment de la saisie des données.

Elles permettent à l'auteur de vérifier s'il a le droit ou non de placer des fichiers en libre accès et le renseignent sur les démarches à entreprendre et les moyens existants pour pouvoir le faire (définition des pré- et post- prints, lien vers les sites détaillant la politique des éditeurs, explications sur l'Open access, possibilités de limitation de l'accès au fichier texte, etc.).

Ainsi le service Infoscience s'est enrichi d'une prestation de conseil en droit d'auteur, assuré par un spécialiste de l'édition scientifique appartenant aux Presses polytechniques universitaires romandes (PPUR), via une Foire Aux Questions (FAQ) disponible sur le site (<http://infoscience.epfl.ch/copyright/>) et un service personnalisé de questions réponses accessible par e-mail. Des séminaires et conférences destinées aux chercheurs sont aussi prévus.

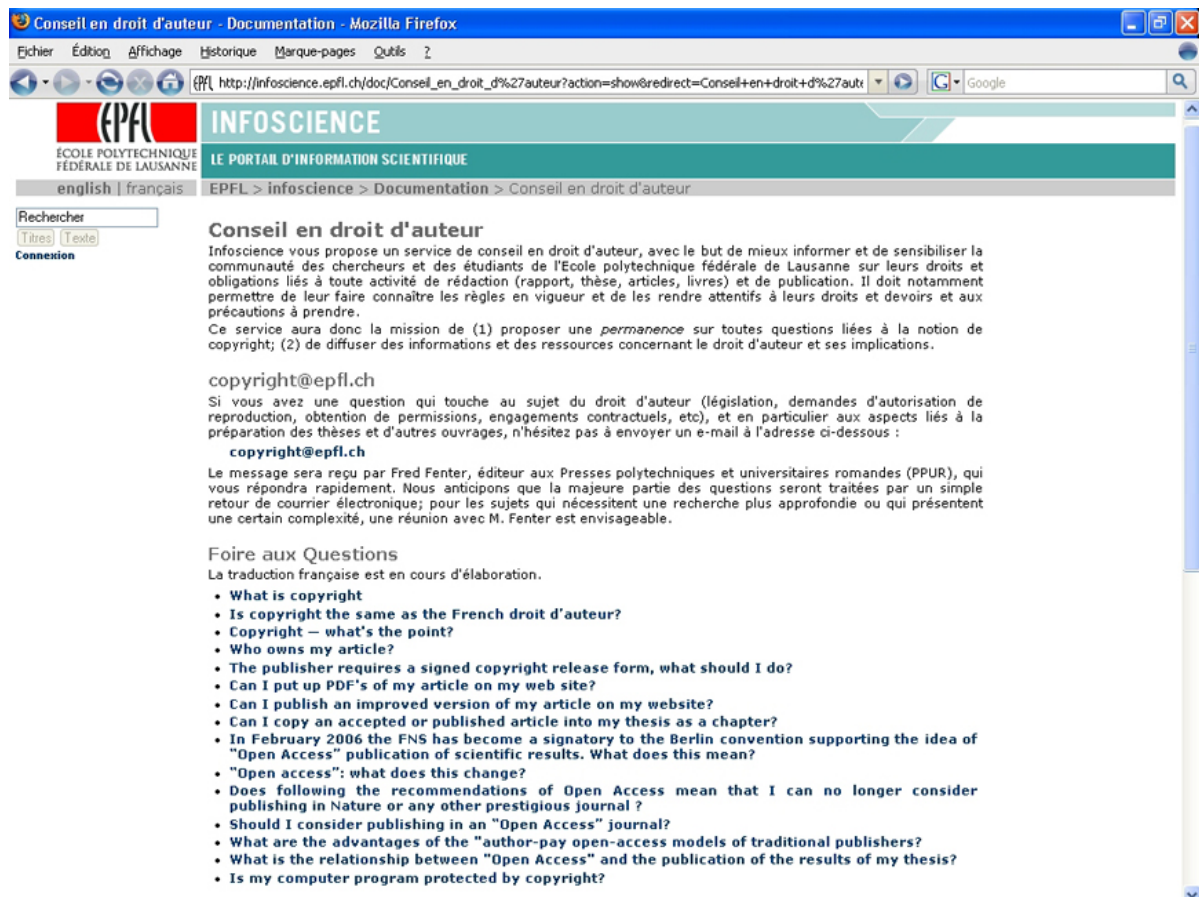


Figure. 6 : Ecran d'accueil du service de conseil en droit d'auteur

Intégration dans le moteur de recherche interne de l'EPFL

L'EPFL a installé un moteur de recherche local (<http://search.epfl.ch>), basé sur la solution Google Search Appliance (<http://www.google.com/enterprise/gsa/index.html>), afin de rendre plus simple l'exploration de ses 300 000 pages web, de ses multiples bases de données publiques et de son annuaire. Infoscience est intégré à ce moteur, de manière très visible (voir la copie d'écran ci-dessous) pour permettre de trouver plus facilement les publications d'un chercheur via cet outil généraliste très utilisé en interne et depuis l'extérieur. C'est aussi un moyen de promotion efficace d'Infoscience.

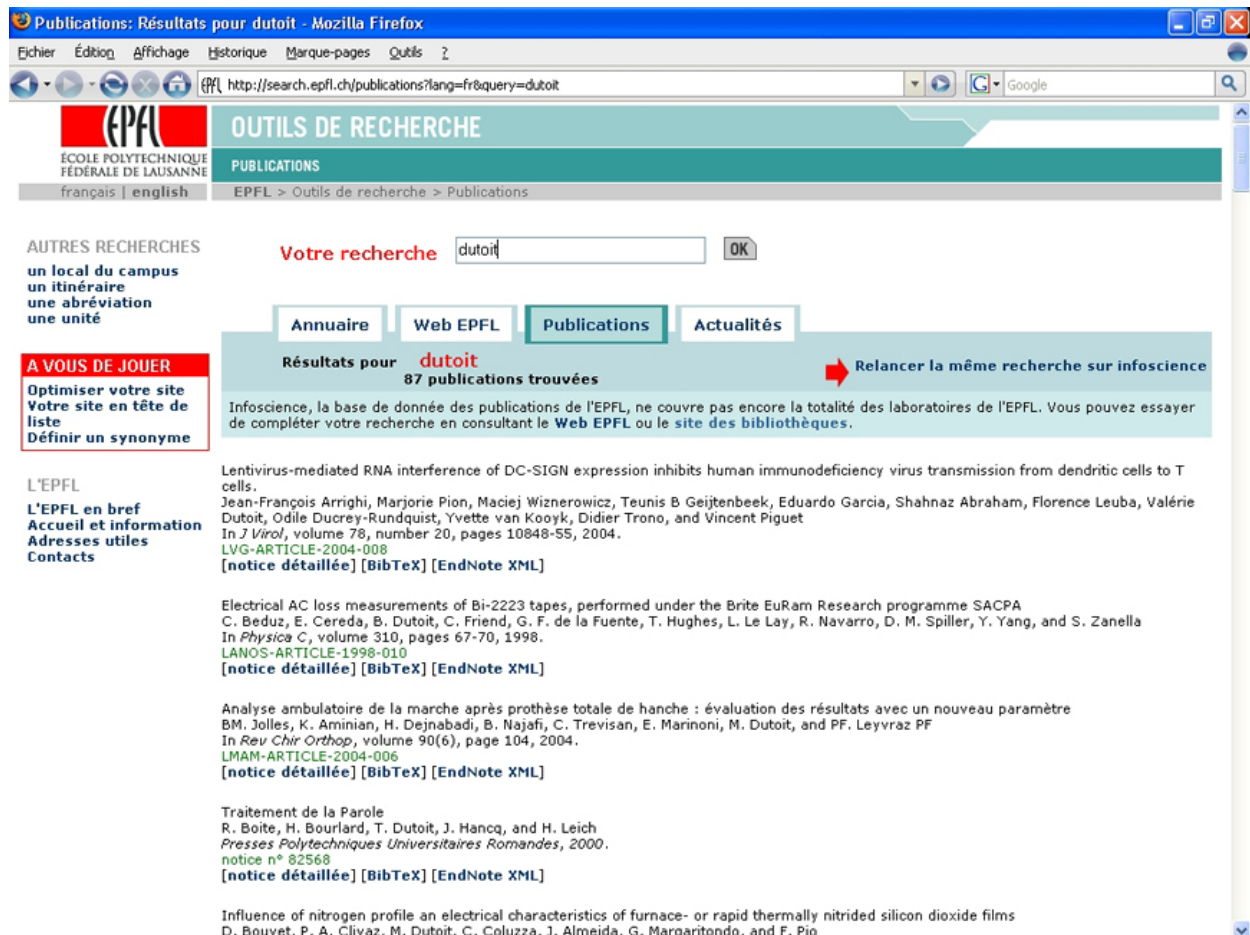


Figure. 7 : Interface du moteur de recherche interne EPFL

Principes de conception. Une archive institutionnelle pour quoi faire ?

Au-delà de la collecte, le service

Infoscience a été conçu comme un *service* de collecte et de valorisation du patrimoine scientifique. Le patrimoine scientifique étant, selon nous, constitué,

- des productions scientifiques de l'EPFL
- des collections de documents scientifiques acquises par l'EPFL
- mais aussi des hommes et des femmes qui font de la recherche à l'EPFL, ces personnes étant décrites par leurs profils personnels mentionnant leurs activités scientifiques et leurs compétences.

Second principe, une archive institutionnelle doit être alimentée par ceux-là mêmes qui créent ce patrimoine, autant parce que cela influe directement sur l'image scientifique qu'ils donnent à leurs pairs et au public (voir plus loin la question de la notoriété), que pour donner à cette activité continue, impliquant de très nombreuses personnes et pratiquée à l'échelle d'une institution académique, la souplesse et la dynamique nécessaires [4].

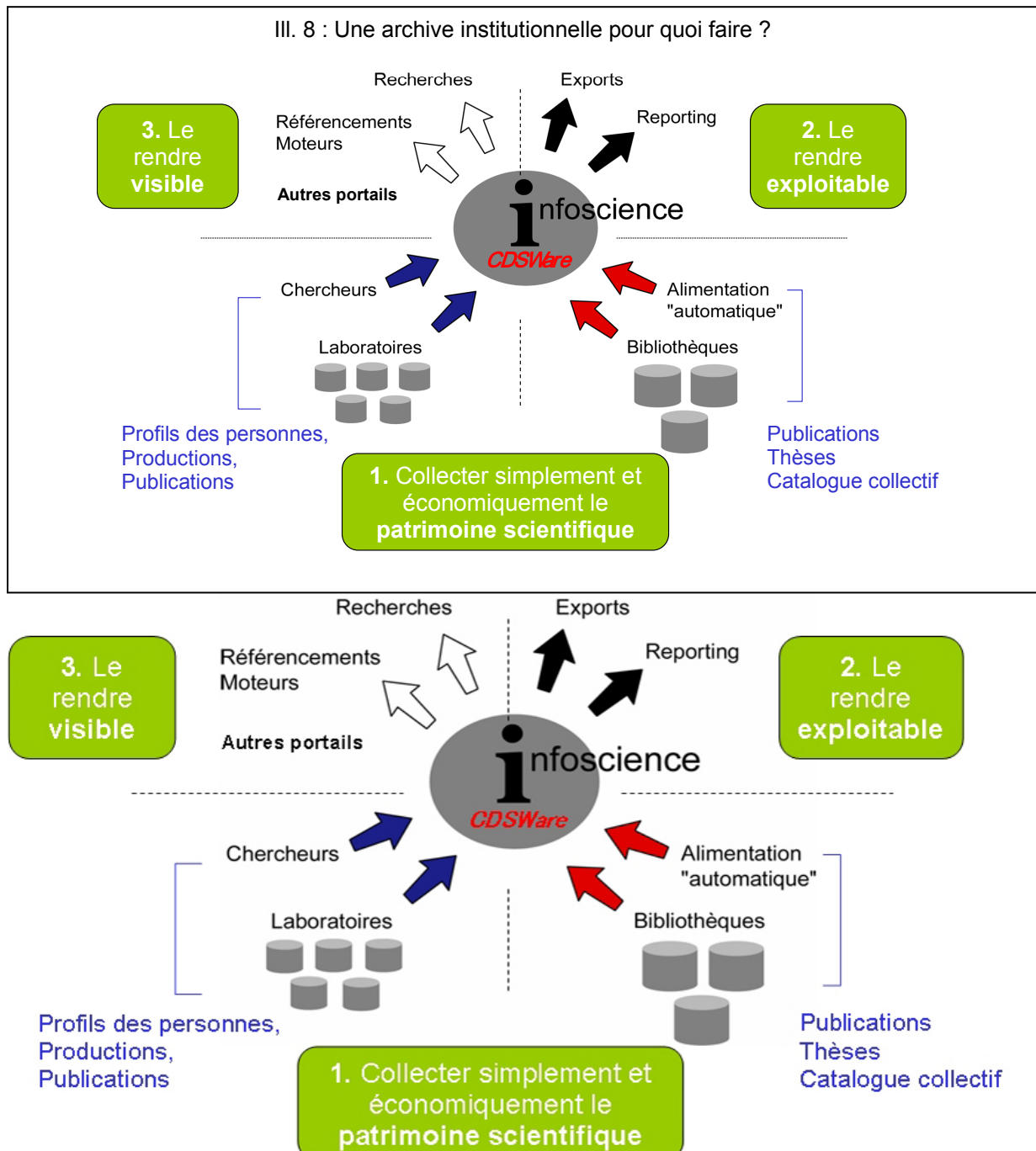


Figure 8 : Une archive institutionnelle pour quoi faire ?

Sa valeur ajoutée vient de ce qu'il répond à des besoins identifiés chez des acteurs et publics de la recherche dans l'Ecole.

L'idée clé est que toute publication saisie une fois dans Infoscience est réutilisable par le chercheur qui l'y a mise, par son équipe de recherche, sa faculté, ou la Direction de l'Ecole. Il n'y aura plus jamais à la ressaisir dans aucune autre application. Et elle sera conservée indéfiniment, avec ou sans full text, et pourra être enrichie au cours du temps.

La réappropriation des publications par les utilisateurs est un service essentiel sur lequel l'équipe est en permanence mobilisée pour répondre aux besoins nouveaux, même individuels. Un guichet infoscience@epfl.ch est ouvert pour enregistrer toutes les demandes d'aide ou d'amélioration. L'état des réponses fournies est géré grâce au logiciel OTRS (<http://otrs.epfl.ch>) servant au suivi du « Help desk » informatique de l'EPFL.

Les utilisateurs d'infoscience sont des « CONSOMMATEURS »

Chercheurs, directeurs de laboratoires, doyens de facultés, membres de la Direction contribuent à Infoscience, en l'alimentant ou en précisant son cahier des charges, parce qu'ils en tirent des bénéfices qui leur simplifient la vie ou même l'améliorent.

Infoscience est un outil de travail pour différentes catégories d'acteur, et aussi un outil de pilotage de l'activité :

Acteur interne	Effet direct	A court terme	A moyen terme
Chercheur	Visibilité	+ de citations	Meilleur IF* dans WoS**
Laboratoire	Visibilité Efficacité	Idem	Idem + économie de moyens + mémoire
Faculté	Visibilité Tableau de bord	Meilleure information	Meilleures décisions
Direction EPFL	Tableau de bord	Meilleure information	Meilleures décisions
	Annalyse des données de publication	Anticipation et réactivité	Renforcement du dialogue avec le Conseil des EPF
	Cohérence des données dans WoS	Fiabilité	Influence sur les classements

*IF : Impact Factor

**WoS : Web of Science, produit par Thomson ISI

Figure 9 : Bénéfices attendus de l'archive institutionnelle

Le contact direct et la personnalisation des contacts avec les chercheurs et laboratoires, suivis du bouche à oreille et complétés de quelques messages diffusés à travers les listes de discussion ou les organes d'information internes ont été nos principaux outils de sensibilisation.

Des discussions nombreuses ont eu lieu avec plusieurs membres de la Direction, afin de cadrer l'étendue du service et de garantir que – grâce à la qualité scientifique du contenu – l'image scientifique de l'EPFL soit réellement valorisée par ce service.

Chercheurs et laboratoires

Chaque laboratoire qui entre dans Infoscience bénéficie d'un service personnalisé pour la reprise de son patrimoine existant (analyse de ses données, import et validation) et l'adaptation du service à ses besoins (site web, formats d'export, ...).

Doyens de facultés

Nous pensons qu'Infoscience doit permettre à un Doyen de faculté d'avoir une vision claire et à jour de la production scientifique des laboratoires qui composent sa faculté. C'est pour lui également un outil facilitant le dialogue avec les directeurs des laboratoires.

Dans l'état actuel du développement d'Infoscience, nous n'avons pas assez de retours de leur part pour savoir s'ils vont utiliser l'outil comme instrument de pilotage.

Direction de l'EPFL

Il est vital pour la Direction de pouvoir disposer de moyens de valoriser l'image scientifique de l'EPFL vers la communauté scientifique internationale comme vers le grand public et de pouvoir exprimer la performance scientifique générale de l'Ecole lors des négociations avec les instances de tutelle.

Infoscience n'est qu'un des moyens de le faire, et ne vise pas en soi à être un outil biblio ou scientométrique. Pour répondre aux besoins de la Direction, nous avons commencé à développer les

capacités de l'outil à signaler les publications de qualité, notamment celles qui sont référencées dans Web of Science, pour lesquelles un facteur d'impact (IF) existe [5] et celles qui sont « peer reviewed » (voir plus loin, le chapitre consacré à la qualité scientifique).

Nous avons en projet la réalisation d'un système de comparaison entre Infoscience et WoS, permettant de vérifier que les publications référencées dans Infoscience et qui devraient l'être également dans WoS s'y trouvent bien. Si cela n'est pas le cas nous en informerions WoS en demandant une correction des données.

Communautés scientifiques, médias et grand public

Comme décrit plus haut dans la partie consacrée aux fonctionnalités et aux services d'Infoscience, l'équipe technique s'occupe d'intégrer l'archive Infoscience dans le web scientifique mondial afin de servir à la communauté scientifique son contenu en information.

Pour ce qui est des médias et du grand public, nous envisageons le développement de services et d'outils spécifiques, basés sur une « boussole scientifique » (moteur de recherche acceptant le langage naturel pour répondre à la question « qui est spécialiste de quoi ? ») et des services d'information de type « push » (agrégateurs de nouvelles, alertes, etc.). Ces services sont encore à l'état de projet.

La prise en compte de la « qualité scientifique » des publications est un facteur important d'acceptation du projet

Dans le monde d'avant le web, la performance des chercheurs pouvait s'analyser au travers du Science citation index, devenu Web of Science. La notoriété était déjà un paramètre important, dépendant des citations, mais également d'autres facteurs, comme l'appartenance à des comités éditoriaux de revues prestigieuses, ou la participation active au collège invisible.

Aujourd'hui, avec le développement du web, la visibilité online, c'est-à-dire la facilité avec laquelle on retrouve un chercheur sur le web et la richesse des informations que l'on trouve sur lui-même et ses activités scientifiques, devient un facteur essentiel de notoriété.

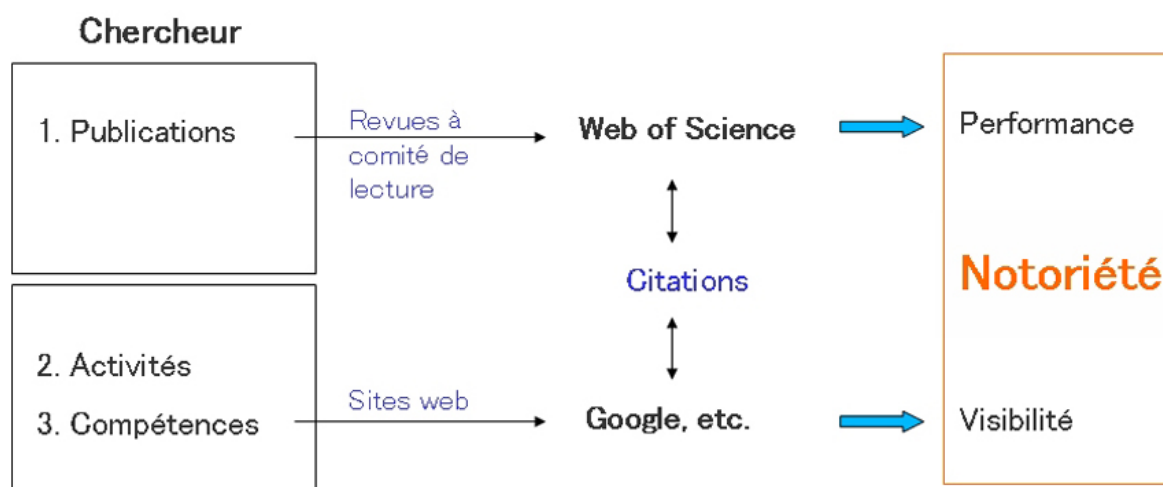


Figure 10 : Quel(s) facteur(s) d'impact aujourd'hui ?

Or les pages web des professeurs ou des laboratoires sont finalement peu visibles car elles sont noyées dans les 10 milliards de pages web constituant un univers d'informations peu structurées et indexé de manière non systématique par les outils de recherche sur le web, même les plus spécialisés (Google Scholar - <http://scholar.google.com/> , Scirus – <http://www.scirus.com> , etc...).

La structuration des données de publication et des profils des personnes qu'apporte une archive institutionnelle couplée à un annuaire évolué permettent alors d'augmenter cette visibilité. Le point critique est celui de son acceptabilité par les chercheurs eux-mêmes, qui doivent être convaincus de la valeur ajoutée de ce type de service, surtout si son alimentation peut être perçue par eux comme une contrainte supplémentaire.

Ce qui distingue peut-être l'EPFL d'autres institutions de recherche, c'est que la compétitivité du monde de la recherche y est reconnue explicitement et que les outils d'influence sont reconnus comme une nécessité. De manière pragmatique, la plupart des chercheurs, mais surtout la Direction de l'École, cherchent à utiliser à leur profit les outils et pratiques de la mesure du mérite scientifique, comme le Web of Science, tout en reconnaissant leurs limites.

Au cours du lancement d'Infoscience, nous avons mesuré à quel point les évolutions récentes liées au développement de l'Open access, auxquelles nous pensions naïvement que chacun allait adhérer sans réserve, suscitaient une certaine incompréhension, voire de la méfiance.

Il nous a fallu prendre en compte la demande de certains chercheurs et des membres de la Direction qui souhaitent que les publications considérées comme étant de qualité soient particulièrement valorisées.

C'est ainsi qu'un consensus a été recherché avec les dirigeants de l'Ecole pour définir les publications admises à apparaître dans Infoscience et pour les catégoriser selon leur niveau de « qualité ».

Publications admises :

- Catégorie A+ : articles signalés dans le citation index de Web of Science
- Catégorie A : articles publiés dans des journaux à comité de lecture, communications de congrès à comité scientifique, documents indexés dans ISI index to Scientific Proceedings Scopus, Inspec, Compendex, Mathscinet, CSA, Medline, Biosis, Avery Index of Architecture...
- Catégorie B : « prépublications, technical reports », articles publiés dans des revues sans comité de lecture, communications de congrès sans comité scientifique, brevets, normes, thèses.

Publications non recommandées :

- La pertinence de ces publications est très variable et elles ne font pas l'objet de peer reviewing scientifique. Exemple : les posters.

Publications non admises :

- Les articles de vulgarisation ou de communication écrits par des chercheurs et publiés dans des revues non scientifiques ne sont pas admis dans Infoscience.

Figure 11 : « Qualité » des publications signalées dans Infoscience. Catégorisation des publications selon le consensus établi au sein du comité de gestion Infoscience

Le risque d'une telle pratique était de diminuer la valeur de l'outil de travail pour le chercheur qu'est Infoscience et de limiter ainsi son adoption à large échelle. En effet si un chercheur ne peut plus placer toutes ses publications dans l'archive institutionnelle, il ne pourra abandonner ses outils personnels de gestion de l'ensemble de ses publications, et n'aura plus d'avantage à adopter l'outil collectif proposé.

Par contre, au niveau de la Direction de l'EPFL, l'archive institutionnelle n'était acceptable que si elle pouvait contribuer à donner une BONNE IMAGE de l'institution et que si elle est pouvait contribuer – ou au moins ne pas nuire – à sa politique de développement de la qualité et de l'influence scientifique de l'EPFL.

Nous avons dans le même temps mis en place un marquage des publications par des indicateurs pouvant signaler (statut de publication : peer reviewed, signalé dans WoS, submitted, accepted, published), qui permet ensuite de donner des **vues** de la base Infoscience selon ces critères. Enfin, la catégorisation qualitative doit maintenant être soumise aux facultés qui auront certainement une vision différente des choses pour certains types de publications, selon les spécificités de leur domaine scientifique. Cette diversité est cependant reconnue et admise par la Direction. Au final, chaque communauté scientifique pourra donc exprimer et voir transcrite dans l'archive institutionnelle sa définition de la qualité scientifique de ses publications.

En conclusion sur ce point, on peut donc dire que l'archive institutionnelle, en se situant au cœur du processus du « publish or perish », est un objet politique sensible. Sa mise en œuvre demande de dialoguer avec les chercheurs et les responsables de l'institution sur ce thème afin que là aussi l'archive apporte une valeur ajoutée ou au moins n'entre pas en contradiction avec les conceptions et les projets des différents acteurs.

L'Open access est un plus... mais pour plus tard

Nous avons détecté au cours du lancement du projet à quel point les chercheurs avaient besoin d'appui pour comprendre et respecter les règles du droit d'auteur dans le contexte nouveau du web généralisé.

Avant même de parler d'Open access [6], il faut souvent répondre à des questions de base comme « à qui appartiennent mes publications ? ».

Par ailleurs, pour plusieurs chercheurs chevronnés avec qui nous avons eu des discussions, et qui sont représentatifs des chercheurs de l'EPFL par leur réputation scientifique élevée, il y a un malentendu vis-à-vis de l'Open access qui, selon eux, en permettant l'auto archivage de publication non validées par les pairs, nuitrait fortement à la qualité de la recherche publiée, alors que le Web of

Science, aussi imparfait soit-il, permettrait – toujours selon leur conception – de ne valoriser que les publications dont la qualité est garantie par le mode de publication et de référencement. D'autres chercheurs, dans des disciplines où les conférences ont beaucoup d'importance, comme l'informatique, ont montré une attitude différente : pour eux, l'auto-archivage des publications sur leur site web leur permettait de développer leur visibilité, sans avoir à trop se soucier de demander l'autorisation à l'éditeur. Par contre le signalement de leurs publications avec les mêmes fichiers full text sur l'archive institutionnelle risquait de rendre tout cela plus visible et donc de provoquer une réaction des éditeurs contre l'EPFL. L'entrée de leur publication dans l'archive signifiait donc pour eux l'obligation de se mettre en conformité avec le droit d'auteur et donc un travail supplémentaire pour l'ensemble de leur publications déjà auto-archivées.

Pour qu'Infoscience tienne compte de ces tensions et de ces pratiques nous avons donc établi par consensus une recommandation concernant les fulltexts et laissé la possibilité aux chercheurs de déposer les fulltext en mode public, ouvert au seul domaine EPFL, ou bien encore privé.

La recommandation envisagée, en complément de la liste des publications admises, prévoit donc ceci :

- « **Fulltext** : Le fulltext des publications est un facteur décisif de visibilité dans les moteurs de recherche de type Google, Google Scholar, Citeseer, etc. Sont acceptées les versions préprints, selon les pratiques des facultés, et postprints, dans le respect de la législation. »

Organisation et missions de l'équipe de projet

Infoscience, c'est une très petite équipe, constituée autour d'un coordinateur technique rattaché au service d'information scientifique et des bibliothèques et d'un architecte en information membre du Service informatique central (DIT : Département IT) de l'EPFL. Le coordinateur technique doit être excellent techniquement, mais surtout doit savoir dialoguer et coopérer avec de nombreux partenaires et utilisateurs (chercheurs, bibliothécaires, spécialistes en droit d'auteur, développeurs du CERN, ...) pour entendre les besoins et les transformer en propositions techniques (développements logiciels, flux de traitement) et de services. Le Directeur de l'information scientifique et des Bibliothèques assume la fonction de chef de projet, et mène le dialogue avec les instances de la Direction pour fixer les objectifs généraux du service. Un comité de gestion Infoscience dirigé par le Vise président pour les Affaires académiques, en charge de la Recherche et de l'Enseignement de l'EPFL, définit les axes stratégiques et valide les étapes définies.

Pour documenter les processus et les retours d'expérience, gérer les développements et les ressources, un Wiki Infoscience (<http://empc51.epfl.ch/infoscience/>) a été ouvert. Il est public en lecture et ouvert à tous les participants de l'équipe de projet en écriture.

Cette équipe est destinée à évoluer, et le plan de développement Infoscience prévoit que le coordinateur va progressivement « rendre les clés » de l'outil aux bibliothécaires scientifiques et aux curateurs dans les facultés, la fonction de support aux utilisateurs et de contrôle des données prenant, avec le temps, plus d'importance que les développements techniques de l'application.

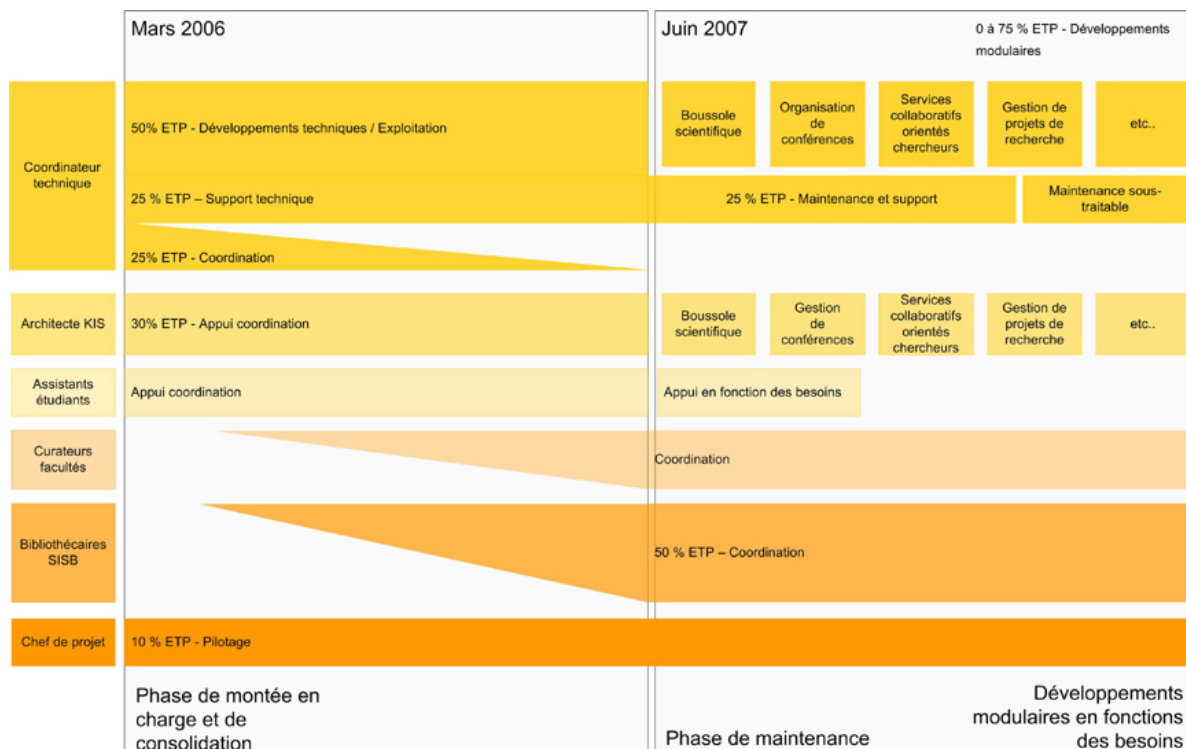


Figure. 12 : Plan de travail et composition de l'équipe de projet, selon le plan de développement

État de développement et perspectives

Aujourd'hui, nous jugeons que le développement d'Infoscience est satisfaisant : 145 laboratoires sur 240 environ ont adhéré **volontairement** au projet et ont décidé d'utiliser Infoscience pour gérer et valoriser leurs publications scientifiques, ainsi que le profil des membres de leurs équipes. De nouveaux laboratoires frappent à la porte chaque semaine.

La croissance du nombre de publications signalées est également un bon indicateur de l'appropriation de l'outil par les laboratoires et les facultés.

III. 12 : Plan de travail et composition de l'équipe de projet, selon le plan de développement

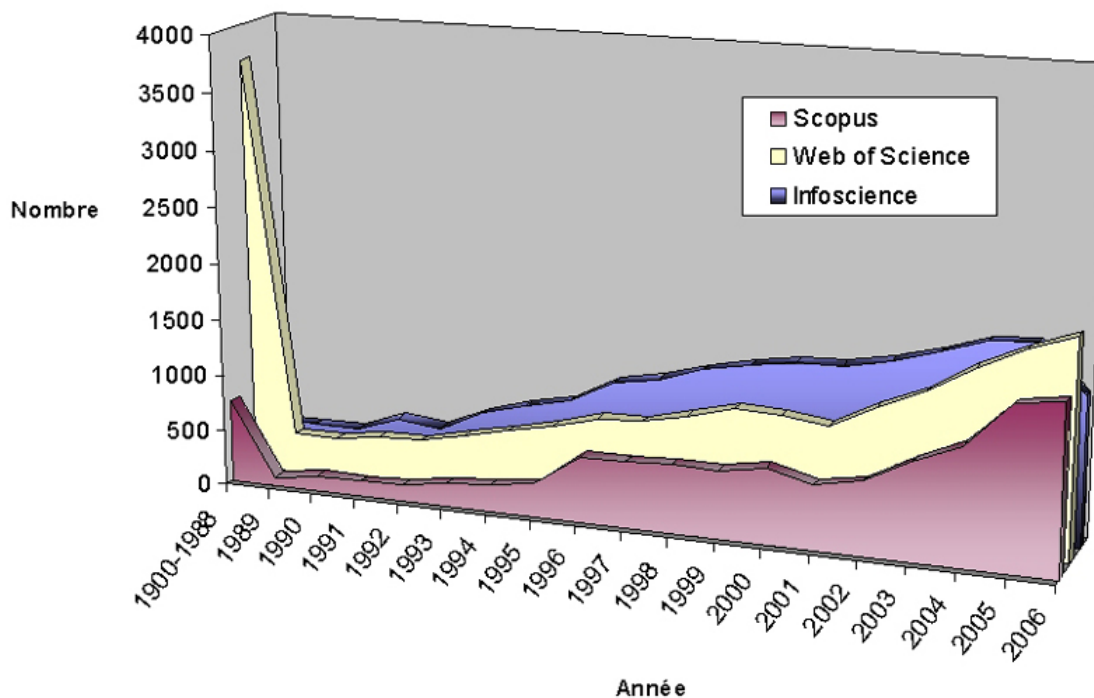


Figure 13 : Progression du nombre de publications signalées dans Infoscience (articles et papiers de conférences), en comparaison avec celles signalées dans web of science et SCOPUS au 15 02 07

Les membres de la Direction de l'École ont donné, début mars 2006, leur soutien au projet et ainsi validé son statut d'activité permanente du service d'Information scientifique et des Bibliothèques de l'EPFL.

Grâce à cette décision de stabiliser l'activité, nous avons maintenant une visibilité de 3 à 4 ans. La Direction a accepté le plan de développement qui prévoit un certain nombre de « délivrables », dont, en substance :

- L'atteinte du nombre maximal de laboratoires participants
- La constitution et la mise à jour de l'annuaire des scientifiques
- La réutilisation des données accumulées pour le rapport scientifique qui a précédé Infoscience
- Le contrôle de qualité et d'exhaustivité des publications de l'EPFL signalées dans Web of Science
- Les outils de contrôle de qualité et d'enrichissement des données

Délivrables	Méthode / Principe	Coût / Moyen
Annuaire des professeurs contenant 90% des profils en français et en anglais, structurés selon modèle commun	Obligation faite à chaque professeur de créer et maintenir son profil	Directive EPLF à tous les professeurs. Adaptation de l'outil people@EPLF pour une structuration plus poussée
Dispositif de contrôle de mises à jour continue de l'annuaire avec garantie d'édition online annuelle à l'échéance souhaitée par la Direction	Détection automatisée des profils non mis à jour	Directive EPLF donnant la responsabilité aux facultés de contrôler la mise à jour des profils. Développement d'un outil de monitoring par l'équipe Infoscience
Liste exhaustive des publications EPLF signalées dans Web of Science 2000-2006	Récupération de la base des publications gérée jusqu'en 2006 par la Délégation à la Recherche de l'EPLF	Travail commun entre le coordinateur technique Infoscience et le service du Délégué à la Recherche
Signalement complet et exact de toutes les publications EPLF dans Web of Science	Outil de contrôle de cohérence Infoscience et WoS. Demande de correction des erreurs WoS	Développement d'outils d'analyse automatique de la correspondance. Vérifications complémentaires des données par les bibliothécaires scientifiques du SISB
Gestion partagée de la qualité des données	Les curateurs de faculté et Bibliothécaires SISB assurent les contrôles, nettoyages et enrichissement de données	Co-développements EPLF et CERN d'outils de contrôle et d'enrichissement semi automatiques (interface « qualité », enrichissement automatique avec Digital Object Identifier (DOI) et le lien vers la version officielle des fulltexts,...)

Figure 14 : Plan de développement Infoscience 2006-2007

Conclusion

Infoscience, et au-delà toute archive institutionnelle, est donc selon notre définition et notre pratique, un service global qui va de l'amont à l'aval dans le soutien à la diffusion de l'information scientifique d'une institution : de la collecte de la production de ses chercheurs, à la visibilité globale sur le web, en passant par le conseil en droit d'auteur et la contribution aux outils de pilotage de la recherche scientifique.

Il ressort de notre expérience que la réussite de la mise en œuvre d'une archive institutionnelle, et notamment son appropriation par les communautés d'utilisateurs scientifiques demande une analyse fine de la culture de l'institution afin de lui proposer un outil acceptable économiquement et politiquement parlant, une maîtrise technique donnant la capacité d'adapter l'outil de gestion aux besoins de la communauté, et enfin une offre de services complémentaires afin de partager le savoir et le savoir faire autour de l'évolution du droit d'auteur et de l'Open access [7]. Pour le Service d'information scientifique et des bibliothèques et le Département IT de l'EPFL, sa mise en place est la

concrétisation d'une mission de soutien aux chercheurs et à la diffusion de l'information scientifique et technique.

[1] CDSware (CERN Document Server Software)
Sitas A, Library Hi Tech 2006, Vol: 24 (3), pp 420 – 429
DOI: 10.1108/07378830610692172.
<http://www.emeraldinsight.com/0737-8831.htm>. [En ligne. Visité le 01 11 2006]

[2] The Complete List: The Top 100 Global Universities
Newsweek international ed., Aug. 13, 2006, Web edition : <http://www.msnbc.msn.com/id/14321230>.
[En ligne. Visité le 01 11 2006]

[3] e-Theses and the OAI e-collection service at EPFL
Aymonin D, Présentation à ELAG 2005 (European Library Automation Group), mercredi 1er juin 2005, CERN, Genève,
<http://indico.cern.ch/sessionDisplay.py?sessionId=82&slotId=0&confId=045#2005-06-01>
[En ligne. Visité le 01 11 2006]

[4] Understanding Faculty to Improve Content Recruitment for Institutional Repositories Foster, NF, Gibbons S. D-Lib Magazine, January 2005, Volume 11 Number 1,
<http://dlib.org/dlib/january05/foster/01foster.html>
[En ligne. Visité le 01 11 2006]

[5] THE ISI IMPACT FACTOR
<http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/>. [En ligne. Visité le 01 11 2006]

[6] What is Open access Bailey, CW Jr. In Jacobs, N, Eds. Open access : Key strategic, technical and Economic Aspects. Chandos.
preprint online, 2006,
<http://www.digital-scholarship.com/cwb/WhatIsOA.pdf> [En ligne. Visité le 01 11 2006]

[7] The Open Research Web: A Preview of the Optimal and the Inevitable
Shadbolt, N., Brody, T., Carr, L. and Harnad, S. (2006) The Open Research Web: A Preview of the Optimal and the Inevitable, in Jacobs, N., Eds. Open Access: Key Strategic, Technical and Economic Aspects, chapter 21. Chandos. <http://eprints.ecs.soton.ac.uk/12369/>
[En ligne. Visité le 01 11 2006]