



Infoscience

ACRONYME	Infoscience
MANDAT	EPFL
ÉTUDIANT	Sylvain Egger
PROFESSEUR-E-S	Omar Abou Khaled, Houda Chabbi Drissi
EXPERT-E-S	Jean-Yves Lemeur, Christine Vanoirbeeck
N°	D07I11
TYPE	Projet de diplôme 2007
CONTACT	Pierre Crevoisier

Infoscience est la plateforme de gestion bibliographique de l'EPFL. Cette plateforme est implémentée sur Invenio. Invenio est un moteur de base de données bibliographique qui offre des outils comme la gestion de collections de publications, un moteur de recherches, l'insertion, la modification et la suppression de publications par les laboratoires de l'EPFL, ...etc. Invenio est une application développée et maintenue par le CERN. Elle est implémentée dans le langage Python.

Un certain nombre d'outils a été développé afin de permettre de maintenir la qualité des données bibliographiques d'Infoscience. Ces outils sont rassemblés sous le nom de "Curator".

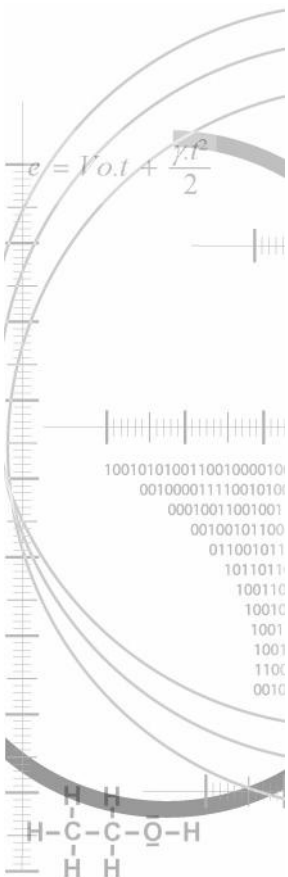
A l'heure actuelle il n'existe que deux types de droits permettant la modification des notices (Dans un catalogue de bibliothèque la notice est un ensemble d'éléments présentant un document : le ou les auteurs, le titre, l'éditeur, la date de publication, le nombre de pages, la collection, ISBN...) :

1. L'administrateur qui permet de modifier n'importe quelle notice de n'importe quel laboratoire.
2. Personnes accréditées au sein d'un laboratoire qui ont des droits que sur ce dernier.

Des fonctionnalités de bases doivent être agrégées au système :

- Uploader un fichier de notices : Mettre à jour des notices en batch en uploadant, à travers l'interface web, un fichier au format XML MARC 21.
- Gérer des doublons : permettre aux curateurs de traiter et supprimer les doublons (un doublon c'est une paire de notice se ressemblant très fortement).
- Archiver un laboratoire : être capable de faire passer son statut d'actif vers archivé.
- Définir le format d'autorité d'un nom d'auteur : choisir le format de nom qui doit être utilisé et associer les formes comme étant des synonymes.
- Transférer les publications d'une personne d'un laboratoire vers un autre.
- Etre capable de modifier le type d'une notice.
- Supprimer une notice d'un laboratoire.
- Ajouter ou modifier un journal scientifique à la base de référence.
- Ajouter ou modifier un livre à la bibliothèque d'un laboratoire.

Ce projet se déroulera au sein de l'EPFL.

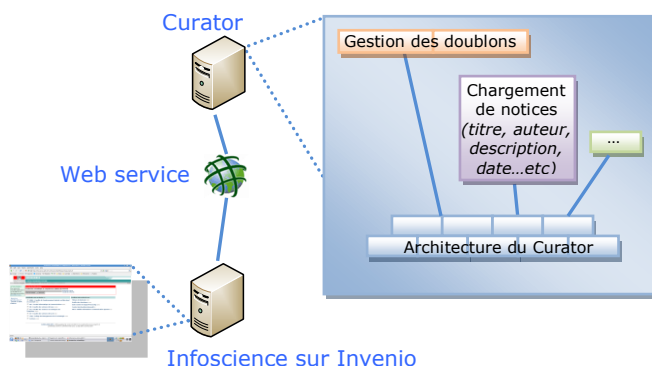


Problématique

L'importance des bibliothèques numériques prend une ampleur très importante, essentiellement dans le monde de la recherche et du développement. L'EPFL, dotée actuellement d'Infoscience¹, souhaite assurer la qualité des informations stockées. Le Curator regroupera tous les outils qui permettront à Infoscience de rester cohérent et efficace.

Objectif du projet

L'objectif du projet n'était pas de réaliser le Curator dans son entier mais de **concevoir et implémenter l'architecture** de celui-ci et d'implémenter le module "Chargement de notices". Cette architecture, schématisé ci-dessous, est le socle du Curator et fournit aux modules les fonctions nécessaires pour agir sur les notices d'Infoscience.



Format des notices

Les données transitent dans un format XML MARC. Il permet de stocker les données bibliographique dans des champs XML tel que le titre, le ou les auteurs, ... etc. Une fois les données XML MARC reçu par le Curator, nous pouvons agir sur celles-ci aisément.

Description générale et technique

Le Curator est un **outil de gestion de qualité** qui agit sur une base de données bibliographique. Il existe différents moteurs de base de données bibliographique, c'est pourquoi l'architecture du **Curator doit être générique** et donc pouvoir se coupler à n'importe quel moteur.

Le schéma suivant démontre le **flux de données entre Curator et Invenio**, Invenio étant le moteur d'Infoscience, la plateforme bibliographique de l'EPFL visible à l'adresse :

- <http://infoscience.epfl.ch>



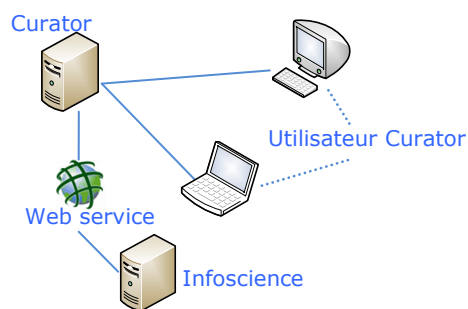
On voit que les données sont extraites d'Infoscience, envoyées au Curator afin d'être traitées. C'est ce traitement qui assure la gestion de la qualité.

Pour traiter ces données, nous avons utiliser Pybliographer, une librairie Python qui offre une API complète pour manipuler les notices importées en XML MARC, provenant d'Invenio.

Le curator est implémenté dans le langage Python. Ce langage est déjà utilisé pour la plateforme Infoscience, ce qui facilite l'interaction entre le Curator et celle-ci.

Communication entre Invenio et Curator

Etant donné que le Curator et Infoscience sont placés sur deux serveur différents, il faut un moyen qui permettent la communication entre ces deux plateformes. Le "moyen de transport" utilisé entre Invenio et Curator est un **web service**. En pleine expansion, les web services fournissent un moyen idéal pour deux applications qui souhaitent communiquer à distance. Ainsi, le Curator pourra envoyer des requêtes à Invenio tel qu'une demande d'authentification de l'utilisateur Curator, des recherches de notices d'Infoscience,... etc.



Plus d'informations

Pour de plus amples informations, vous pouvez me contacter à l'adresse suivante :

Sylvain.Egger@gmail.com

Mots-clés : Python, Infoscience, Curator, gestion bibliographique, MARC XML, Webservice

¹ Infoscience est la plateforme de gestion bibliographique de l'EPFL.