

Projet international
Global Biodiversity Information Facility
<http://www.gbif.org>

Réalisation du nœud informatique de GBIF-Suisse
<http://www.gbif.ch>

Travaux réalisés 2004-2007 et perspectives 2008-2011

F. Burri, M. Bouzelboudjen

Université de Neuchâtel
Service Informatique et télématique

1. But du document

Ce document présente les travaux effectués pendant la période 2004-2007 ainsi que les perspectives de développement 2008-2011 s'inscrivant dans le cadre du projet de nœud informatique de GBIF Suisse. Les réalisations 2004-2005 sont présentées dans le rapport: « Projet international GBIF : Etat de situation au 20 février 2006 ».

2. Introduction

Le nœud informatique a pour mandat d'offrir une plate-forme nationale assurant la fiabilité et la pérennité des flux d'informations. Dans ce cadre il développe et administre des bases de données accessibles via Internet dans les domaines de la faune principalement et de la flore (spécimens des collections scientifiques des Musées de Suisse).

L'Université de Neuchâtel réalise le nœud informatique de GBIF Suisse et met en plus à disposition des chercheurs, étudiants et du public un réseau de compétences et de connaissances dans le domaine de la biodiversité.

3. Réalisations 2004-2007

3.1. Travaux réalisées en 2004

L'architecture informatique déployée dans le cadre du projet GBIF est décrite dans l'exposé du 28 janvier 2004 présenté à Berne. Le nœud informatique du « participant node » suisse est déployé sur une architecture multi-tiers. Cette architecture est composée d'un serveur d'application dans lequel sont installés les logiciels spécifiques au projet GBIF. Le serveur d'application constitue l'interlocuteur immédiat des internautes. Le serveur de base de données s'appuie sur une architecture multi-processeurs connecté à une unité de stockage de masse de type SAN. Les données seront gérées par le SGBD Oracle. Le serveur de base de données concrétise le concept de « data nodes » centralisés. Il contient les données de toutes les institutions qui ont choisi cette solution pour leur intégration

dans le cadre du projet GBIF Suisse. C'est ainsi que deux alternatives ont été proposées aux institutions qui souhaitent adhérer à la logique du projet International GBIF:

- Intégration de leurs données par le noeud centralisé (serveurs GBIF-CH) ;
- Intégration de leurs données dans un serveur propre à l'institution.

Dans sa version de test, la base de données Oracle 9i et l'application cohabitent sur le même serveur. Ce dernier est équipé d'un processeur de type Xeon sous Linux/Redhat. Il dispose d'une capacité de stockage de 74 Gb et d'une mémoire RAM de 4 Gb.

En septembre 2004, la solution décrite a été réalisée et a été présentée à Berne. Cette architecture s'articulait autour des logiciels standards fournis par GBIF International permettant l'intégration du format Darwin Core.

En résumé, les tâches suivantes ont ainsi été réalisées :

- Achat et installation d'un serveur sous Linux ;
- Installation d'une base Oracle 9i ;
- Intégration de données de test issues du musée d'histoire naturelle de Neuchâtel ;
- Installation des packages standards GBIF (Digir Provider, GBIF Portal Toolkit, GBIF Data repository Tools, DiGIR Portal) ;
- Intégration de données test au format texte et MS Excel ®.

Depuis décembre 2004, le comité scientifique GBIF Suisse a décidé d'intégrer dans la base de données, des données de type image. Cette décision est importante. Ainsi, le format de données initial XML Darwin-Core est abandonné au profit du format ABCD (projet BioCASE). Les conséquences de cette décision sont :

- Abandon des outils DiGiR ;
- Abandon des applications liées à DIGIR réalisées par le noeud informatique GBIF-CH à l'Université de Neuchâtel ;
- Installation des nouveaux outils BioCase dans le serveur de test Norma.

3.2. Travaux réalisés durant l'année 2005

Le portail permettant la publication des informations en rapport avec GBIF Suisse (<http://www.gbif.ch>) a été ouvert le 31 mai 2005. Ce site a été développé avec l'environnement Jahia. (<http://www.jahia.org>). Jahia est un CMS (Content Management System) ou système de gestion des contenus.

Pour la mise à disposition des données suisses de biodiversité accessibles via Internet, les actions suivantes ont été réalisées :

- Consolidation de l'architecture de la base de données du noeud Suisse
- Intégration de données tests
- Installation et configuration des logiciels nécessaires (PyWrapper) à la publication de ces données au format ABCD de BioCASE
- Enregistrement des sources de données auprès de GBIF International

La solution informatique basée sur BioCASE a été présentée à la commission scientifique, le 31 août 2005 à Berne.

En septembre 2005, un concept de backup Oracle ® de la base de données test GBIF-CH a été élaboré et mis en oeuvre. Un document présentant la solution a été mis à disposition.

En octobre 2005, un document décrivant la procédure de transmission des flux d'information au nœud informatique central suisse (GBIF-CH) a été rédigé et mis à disposition. Ce document détaille les trois volets suivants :

- Information descriptive de l'institution ;
- Information descriptive de chacune des bases de données fournies par l'institution ;
- Données de collection aux formats MS Access ® ou MS Excel ®.

De plus, ce document précise le rôle du coordinateur GBIF. Ce dernier est le correspondant entre les institutions et le nœud informatique GBIF Suisse. Il doit être informé de l'existence des données et assure la « viabilisation » de ces données pour les nœuds GBIF Suisse.

3.3. Travaux réalisés durant l'année 2006

La réalisation du nœud informatique suisse a nécessité le développement de méthodes et d'outils indispensables pour répondre aux défis que représente l'intégration des données issues des différentes cultures linguistiques Suisse dans le réseau international GBIF. Ces efforts se sont traduits par le développement et la mise en oeuvre d'outils logiciels à même de réaliser des opérations de validation et d'enrichissement des données. Ces outils logiciels réalisés à l'Université de Neuchâtel ont été présentés lors de la journée d'information GBIF-Suisse du 8 novembre 2006 à Berne. Ces développements ont été acceptés par le comité scientifique de GBIF-Suisse. Le but de ces développements est de permettre de :

- Réduire au maximum l'effort qui doit être fourni par les institutions pour publier leurs propres données sur le réseau international de GBIF ;
- Garantir un niveau de qualité des données conforme aux exigences fixées par le comité scientifique de GBIF-Suisse ;

Pour répondre à ces exigences, la base de données du nœud informatique suisse a été construite sur la base d'un thésaurus hiérarchique multi-langues. Ce thésaurus porte actuellement sur 14 des 49 concepts (colonnes) prévus dans la base de données. Il offre quatre fonctionnalités essentielles :

- Validation du contenu ;
- Correction automatique ;
- Enrichissement des données ;
- Traduction automatique vers une langue ou sur une valeur de référence.

En mai 2006, les premières données à caractère botanique (« Collection Data on the herbarium and living plants of the CJBG ») fournies par le « Conservatoire et Jardin botaniques de la Ville de Genève » ont été intégrées dans la base de données du nœud informatique Suisse. Elles ont été publiées sur le portail International de GBIF. La solution d'intégration a fait l'objet d'une présentation au comité scientifique de GBIF Suisse lors de sa séance du 18 mai 2006.

En novembre 2006, les premières données à caractère zoologique fournies par le « Naturhistorisches Museum Bern » ont été intégrées dans la base de données du nœud informatique Suisse. Elles ont été publiées sur le portail International de GBIF. La solution a été présentée lors de la journée d'information GBIF-Suisse du 8 novembre 2006 à Berne.

De plus, la publication des images contenues dans ce lot d'observations des premières données à caractère zoologique a nécessité le développement, sur le nœud informatique suisse, d'un portail Internet capable de formater et d'afficher ces images.

3.4. Travaux réalisés en 2007

3.4.1. Intégration des collections de parasites

Une étude visant à permettre l'intégration des collections de parasites est en voie de finalisation. Les collections concernant les observations de parasites font généralement références à des observations concernant l'espèce hôte de ces parasites. Dans ce contexte, le comité scientifique de GBIF a décidé que cette information devait être exploitée. Cette nouvelle fonctionnalité implique une modification de la structure de la base de données ainsi que le développement des outils nécessaires pour l'intégration de ces données.

3.4.2. Développement d'un portail Internet national de service

Les fonctionnalités offertes par le noeud international de GBIF ne permettent pas d'exploiter les fonctionnalités offertes par le thésaurus hiérarchique multi-langues de la base de données du noeud informatique suisse. La version test de ce portail a été présentée lors de la séance du comité scientifique de GBIF du 20 mars 2007. Ce portail permet d'offrir des services à haute valeur ajoutée, non seulement aux institutions qui adhèrent à la logique du projet GBIF Suisse, mais aussi à l'ensemble de la communauté des chercheurs et des scientifiques. Parmi les fonctionnalités offertes citons les possibilités :

- De réaliser des interrogations multi-langues sur les données ;
- De mettre le contenu du thésaurus à disposition des internautes.

L'ouverture du thésaurus permettra aux internautes de proposer son enrichissement.

3.4.3. Mise à disposition des données dans le réseau HerpNet

A la demande du comité scientifique de GBIF-Suisse, la mise à disposition des données dans le réseau [HerpNet](#) a nécessité l'installation du logiciel DiGIR. Des tests d'interrogation des données sont actuellement en cours. A l'issue de ces tests, les données disponibles sur le noeud information Suisse de GBIF pourront être prochainement mises à disposition sur le réseau HerpNet en complément au réseau GBIF.

4. Perspectives 2008-2011

En 2008, l'architecture du noeud informatique suisse sera consolidée pour permettre l'intégration des collections de parasites sur la base de l'étude réalisée en 2007. Cela signifie que la base de données GBIF-Suisse ainsi que le portail d'interrogation de ces données vont évoluer afin d'exploiter de manière optimale les données mises à disposition par les partenaires nationaux. L'année 2008 marquera aussi le début de l'intégration massive des données sur le portail national.

De 2009 à 2011, l'intégration des données des partenaires nationaux sera poursuivie. Au regard des informations disponibles à l'heure actuelle, on peut penser que des efforts seront nécessaires pour intégrer de manière optimale les données de certaines institutions.

Dans ce sens, le noeud informatique suisse propose d'initier un projet pilote qui permettra d'évaluer les possibilités offertes par les concepts multi-langues associés à la base de données du noeud informatique Suisse de GBIF à la base de données nationale du Centre Suisse de Cartographie de la Faune (CSCF). Ce projet nécessitera des adaptations de la base de données du CSCF actuel ainsi que la création de passerelles entre les 2 bases GBIF et CSCF.

Enfin, comme pour chaque année, une veille technologique est assurée. On peut prévoir que l'évolution des technologies d'interconnexion entre les bases de données distribuées via le web va conduire le nœud informatique suisse à mettre en place de nouvelles solutions proposées par GBIF international (exemple TAPIR) qui remplaceront les protocoles actuels (BioCase, DiGIR).

Neuchâtel, le 4 décembre 2007

NB :

Les documents cités dans le texte sont disponibles à l'adresse : <http://www.gbif.ch/page6215.html>

Service informatique et télématique (SITEL)

■ Rue Emile-Argand 11 ■ CP 2 ■ CH-2007 Neuchâtel 1
■ Téléphone : +41 32 718 20 00 ■ Fax : +41 32 718 20 01 ■ E-mail : secretariat.sitel@unine.ch ■ www.unine.ch/sitel
