

# Projets liés aux bases de données botaniques

Par Reeb Catherine, 2002

Un essai de synthèse des projets liés aux bases de données botaniques

Catherine REEB, laboratoire Informatique et Systématique, UPMC, 12 rue Cuvier, 75005 Paris  
[catherine.reeb@snv.jussieu.fr](mailto:catherine.reeb@snv.jussieu.fr)

## Contexte

### Introduction

L'émergence des projets de gestion des données botaniques est intimement liée à la réflexion globale sur la diversité des espèces et sur la nécessité de sa gestion raisonnable et raisonnée. La notion de « biodiversité » englobe la richesse du vivant à toutes les échelles d'appréhension : échelles de la population, de l'espèce, de l'écosystème, recouvrant respectivement les domaines de la génétique, de la systématique et de l'écologie.

Parallèlement la technologie informatique a évolué très rapidement et des réflexions sont menées depuis une vingtaine d'années sur la structuration des connaissances, leur représentation informatique et leur analyse rapide et poussée permise par la puissance des algorithmes et la rapidité croissante des processeurs. Ces projets se situent donc dans un contexte historique récent que nous survolerons dans un premier temps.

Le but de cette synthèse est surtout de comprendre les objectifs, les initiateurs et les acteurs, les moyens mis en oeuvre par ces projets internationaux et nationaux liés aux bases de données botaniques, ainsi que leurs éventuels liens les uns avec les autres.

## Cadre historique

### Historique succinct des liens informatique-systématique (Bailly, 2001)

**1960** : Premiers essais d'informatisation d'une collection

- **fin années 60** : premiers essais de génération de clé de détermination automatique
- **1978** : parution du premier ouvrage synthétique sur les problèmes liés à l'identification en biologie : "Biological identification. The principle and practice of identification methods in biology" (Pankhurst, 1978).
- **années 1980** : développement des langages dédiés à la représentation des connaissances taxonomiques : DELTA(1980) <http://biodiversity.uno.edu/delta/>, X-PER (1983) <http://lis.snv.jussieu.fr/Xper/>.
- **Le développement du réseau Internet et du World Wide Web**, rendant accessible au grand public depuis une dizaine d'années un flux et une masse d'informations énorme ne peut que favoriser le développement de la diffusion des connaissances.

- **En France : 1995**, la réunion annuelle de la société française de systématique a pour thème « informatique et systématique ». Elle est suivie par la parution du numéro 14 de la revue « Biosystema » sur ce thème « informatique et systématique » (Lebbe, 1996).

### **En parallèle, émergence des projets autour du rassemblement des connaissances et de la protection de la biodiversité.**

- **1985** : TDWG , taxonomic databases working group.
- **Juin 1992** : sommet de la terre à Rio, préparé par l'*UNEP* (United Nations Environment Programme <http://www.unep-wcmc.org> ) établissement notamment de la **Convention sur la Biodiversité** (*CBD* Convention on Biological Diversity ou *CDB* en français, Convention sur la Diversité Biologique) (<http://www.biodiv.org>) dont les trois objectifs fondateurs sont : la **conservation de la biodiversité, l'usage raisonnable de ses composants et le partage équitable des bénéfices issus de son utilisation**. Au delà de ces principes fondateurs, les gouvernements sont sollicités pour la mise en place et le développement, au niveau de leur pays ou des groupes de pays comme l'Europe, de projets et d'actions leur donnant les moyens de tendre vers et d'atteindre ( ?) ces objectifs. Une dizaine d'engagements plus précis sont inclus dans la convention, dont le premier concerne : « l'identification et la surveillance des composants importants de la diversité biologique qu'il est nécessaire de conserver et d'user avec parcimonie ».

La **COP** (Conference of the Parties ou *CDP*, Conférence des Parties en Français), est l'autorité suprême de la *CBD*. Elle regroupe des représentants de tous les gouvernements ainsi que de nombreux organismes nationaux et régionaux qui ont ratifié le traité.

Le **Clearing House Mechanism** est un programme hébergé sur Internet visant à promouvoir les coopérations scientifiques et techniques ainsi que les échanges d'informations. (voir *Species 2000* par exemple)

Les projets initiés aux niveaux nationaux et répondant aux objectifs de la Convention sur la Biodiversité sont proposés puis évalués via le *GEF* (Global Environment Facility <http://www.undp.org/gef> ) et cofinancés alors par l'*UNEP* (United Nations Environmental Program), l'*UNDP* (United Nations Development programme <http://www.undp.org> ) et la World Bank (<http://www.worldbank.org> ) pour les pays en voie de développement).

- à partir de 1993 :
  - création et mise en place de **nouvelles structures**, nationales et internationales chargées de mettre en place des projets répondant aux objectifs de la Convention sur la Biodiversité (voir *GBIF*, onglet "Projets internationaux")
  - **mise en oeuvre de différents programmes** sur la gestion des données disponibles sur la biodiversité, dont ceux impliqués dans la gestion des données botaniques seront évoqués dans ce document (*Species 2000, Euro+Med, ENHSIN, RefTax...*).

## **Les objectifs généraux de ces projets**

Quelle que soit leur origine et la structure qui les porte, tous ces projets tendent vers les mêmes objectifs : (Bailly, 2001)

- **Stocker et organiser les connaissances** sur les espèces vivantes afin de
- **Les disséminer et permettre leur interrogation**
- **Les actualiser**, en comparant par exemple les données issues de sources différentes et sans relations préalables.
- **Les exploiter**, en créant par exemple des phylogénies à partir des données fournies, des analyses morphométriques...

## De quelles connaissances parle-t-on ?

Plusieurs catégories de connaissances sur les organismes ou groupes d'organismes, par la quantité importante de données disponibles et leur utilité pour l'expertise scientifique, peuvent répondre à ces objectifs :

1. **Les inventaires**, catalogues des groupes taxonomiques rencontrés au sein d'une zone géographique délimitée.
2. **Les données sur la biologie des espèces**, au sens le plus large :
  - Les **descriptions morpho-anatomiques** et les **données écologiques**, dont l'informatisation structurée est à la base de l'I.A.O (identification assistée par ordinateur). L'I.A.O regroupe toutes les applications possibles à partir d'une base de données systématiques (ou base de connaissance), comme :
    - l'implémentation de serveurs d'identification on-line (<http://lis.snv.jussieu.fr>, <http://www.nybg.org/bsci/bfna/> ...) ou la diffusion de CD-ROM interactifs.
    - la génération automatique de clés de détermination, « à la demande », par exemple en fonction du groupe taxonomique étudié ou du lieu de prospection est un champ d'application en plein développement.
  - les **données sur le génome** des espèces.
3. **Les données nomenclaturales**, régies par le code *ICBN* (International Code of Botanical Nomenclature né en 1867 à Paris, dernière édition en 2000 à St Louis, U.S.A) pour le monde végétal et les champignons. Des problèmes de synonymie surgissent si différents noms sont utilisés pour décrire la même entité biologique. Il est nécessaire de connaître les relations entre tous les noms utilisés au cours de l'histoire (...démarrant pour nous en 1753, avec le *Species Plantarum* de Linné) afin de pouvoir utiliser les données sur le taxon étudié sans confusion ni doute possible sur l'identité de celui-ci !
4. **Les données sur les collections disponibles**, les herbiers pour la botanique.

Des liens étroits existent entre ces différents types de données, les informations de terrain ou les informations sur les herbiers n'ont de valeur que si le taxon sur lequel elles portent est clairement identifié, et que les données nomenclaturales complètes et à jour permettent d'accéder à l'information en tenant compte de tous les noms synonymes.

## Les outils disponibles, problèmes soulevés

Tous les projets que nous allons évoquer cherchent à organiser des bases de données existantes, souvent en utilisant un SGBD ou système de gestion de base de données. Une confusion est souvent effectuée entre « la base de données », ensemble structuré de données et le SGBD, qui est le logiciel

permettant de gérer ces données c'est à dire les stocker, d'en permettre la consultation (l'interrogation) et la mise à jour.

La **construction d'une base de données** répond à quelques règles simples que nous ne développerons pas ici (Gardarin, 1999, Stajano, 1998, Delobel et al., 1991). Citons simplement la **règle de non-redondance des données** : une information ne doit être stockée qu'une fois (ex : un nom d'auteur ne doit être écrit qu'une fois dans la base de donnée), **l'interdiction de champs dits « fourre-tout »** contenant des données exprimant des idées différentes, la nécessité de **l'identification unique** de chaque enregistrement...

Le **choix du support** est un problème complexe sur lequel là encore nous ne nous étendrons pas. Très souvent, chaque projet, chaque institut possède son propre système de gestion de données, certains portés par des logiciels commerciaux comme Microsoft Access ou Filemaker. Des efforts de développement de logiciels publics ou libres, dont le code source (lignes de programme rédigées dans un langage informatique permettant le fonctionnement de l'application) est accessible à tous, ce qui n'est pas le cas des logiciels commerciaux, sont actuellement menés par de nombreux informaticiens (voir la première « journée du logiciel libre en Languedoc-Roussillon, *JLM 2001* » diffusé sur tela-botanica, tb-informatique).

La **standardisation de l'expression des données** permettant des échanges sans problèmes entre tous les programmes est l'objectif majeur de certains projets ou groupes (voir le *GBIF*, onglet "Projets internationaux"). Le développement d'outils permettant l'interopérabilité entre bases de données non standardisées est également une préoccupation importante des projets en cours.

Le **problème de la redondance des informations** entre les différents programmes est également un souci : il est inutile de construire à plusieurs reprises les mêmes bases sur les mêmes données. La coopération entre tous les projets et toutes les bases disponibles est en général souhaitée, ce qui peut ajouter à la confusion entre leurs origines, leur site de consultation etc...

## Organismes internationaux

La plupart des organismes cités (sauf le *TDWG*, existant depuis 85 et *Optima*), sont nés de la conférence de Rio, de la ratification de la *CBD* et de la volonté de mettre en application ses recommandations. Ils sont ici présentés dans l'ordre chronologique de leur création.

### **TDWG : Taxa Database Working Group**

<http://www.tdwg.org>

**Origine** : un des premiers groupes de réflexion sur les bases de données taxonomiques, crée en septembre 1985 à l'initiative d'un groupe de botanistes appartenant à différentes structures institutionnelles (conservatoires botaniques et muséums) afin de répondre aux problèmes émergents de la gestion des données nomenclaturales, d'herbiers et des données biologiques. Il porta pendant un temps le nom de « Taxonomic Databases Working Group for Plant Sciences », traduisant son origine, mais en 1994 il lui fut demandé par l'*IUBS* (International Union for Biological Sciences) de ne plus se limiter aux plantes et de travailler sur l'ensemble des groupes taxonomiques; seule la première partie du nom fut alors conservée.

**Objectifs :** l'activité principale du *TDWG* est de développer et de promouvoir des standards pour l'enregistrement et l'échange de données taxonomiques sur les organismes.

**Membres et mandataires :** ce groupement, contrairement aux deux suivants est ouvert à tous, individuels (membres individuels), institutions et groupements autour de projets de base de données (membres institutionnels).

**Moyens :** lors de réunions annuelles, ces standards sont discutés avant d'être ou non approuvés. Ils sont alors publiés afin d'être utilisés le plus largement dans le monde. Parmi les standards reconnus par le *TDWG*, on trouve par exemple la référence des noms d'auteurs botaniques (« Author of plant names » de Brummitt et Powell (Brummitt and Powell, 1992)), les formats et protocoles d'échanges entre herbiers (Conn, 1996), un schéma pour l'enregistrement de la distribution des plantes, le langage *DELTA* (description language for Taxonomy) etc...

## **GTI : Global Taxonomy Initiative**

<http://www.biodiv.org/programmes/cross-cutting/taxonomy>

**Origine et membres participants :** groupement créé dans la dynamique de la *CBD*, en février 98 à Darwin en Australie, lors de la quatrième réunion de la Conférence des parties (la *COP*, autorité suprême de la *CBD*).

**Objectifs :** ce groupement cherche à évaluer la connaissance taxonomique et surtout les besoins des différents pays pour améliorer cette connaissance ainsi que pour mener des actions pour la protection de la biodiversité.

**Moyens :** le *GTI* a édicté un ensemble de recommandations pour la définition des stratégies et des projets nationaux et régionaux. Il met notamment l'accent sur les inventaires, l'utilisation et la mise en valeur des collections, la nécessité d'estimer l'état des connaissances taxonomiques pour chaque état et région...

Le *GTI* est donc un organisme intermédiaire entre la Convention sur la biodiversité et les pays signataires. Il intervient au niveau de l'évaluation de la connaissance, des conseils généraux avant la mise en place d'une stratégie et l'élaboration de projets.

## **IOPI : International Organization for Plant Information**

<http://iopi.csu.edu.au/iopi>

**Origine :** *IOPI* est une commission de l'*IUBS* (International Union of Biological Science). Ce groupement a été créé en septembre 1991, en Australie, à l'initiative de l'*IUBS*. Quarante neuf botanistes de onze pays ont participé à cette première réunion.

**Objectifs :** cette organisation soutient le développement de plusieurs projets coopératifs autour de la création et de la gestion de bases de données taxonomiques sur les plantes, dont les projets « Species Plantarum Project », « Global Plant Check-list », « Database of Plant Databases », que nous développerons dans le paragraphe suivant.

**Membres :** individuels et institutions peuvent intégrer cette organisation.

**Liens avec d'autres organismes ou projets :** l'*IOP*I est lié avec le *TDWG*, est un membre participant de *Species2000*, l'*IAPT* (International Association for Plant Taxonomy) et le *CODATA* (Committee on Data for Science and Technology). Bien que créée avant la ratification de la Convention sur la biodiversité, l'*IOP*I cherche à répondre à ses objectifs ainsi qu'à ceux proposés par le *GTI* et le *GBIF*.

## **OPTIMA : Organisation for the Phyto-Taxonomic Investigation of the Mediterranean Area**

<http://www.bgbm.fu-Berlin.de/OPTIMA/ITN/recommendations.htm>

**Origine:** *OPTIMA* est une association de botanistes travaillant autour de la Méditerranée accueillant individuels et institutionnels.

**Objectifs :** cette organisation est gérée par des conseils et des commissions sur tous les thèmes de la biologie autour des plantes méditerranéennes. La commission « for Information Transfer and Networking » s'est fixé comme objectifs d'établir des bases de données sur les taxons ou les groupements phytosociologiques méditerranéens, d'identifier les projets et les experts informatiques autour de ce thème.

**Liens avec d'autres organismes et programmes :** la *Medcheck-list* a été initiée par *OPTIMA* semble-t-il, mais le projet a été suspendu en 1995 lors de la réunion *OPTIMA* de Séville. Ce projet est maintenant inclus dans l'*EMP* (voir onglet "Projets internationaux").

## **Projets internationaux**

Comme la création des organismes précédents, tous les projets ci-dessous ont été motivés par l'application de la convention sur la biodiversité, la *CBD*, en s'inspirant ou en relation avec les recommandations édictées par le *GTI*.

## **GBIF : Global Biodiversity Information Facility**

<http://www.gbif.org>

**Origine :** l'idée de ce programme a émergé lors de la réunion du groupe de travail « informatique et biologie » de l'*OECD* (organisation for economic cooperation and development <http://www.oecd.org> ou *OCDE* en français) en janvier 1996. Les propositions sont avalisées les 22-23 juin 1999 à Paris et les représentants des pays participants (28 pour le moment) se réunissent régulièrement depuis (3 fois en 2001 par exemple).

**Objectifs :** le but général du *GBIF* est de « promouvoir, coordonner, construire, implémenter la compilation, les liens, la standardisation et la dissémination des données sur la biodiversité mondiale ». L'échelle d'appréhension est donc mondiale, bien que les données puissent être traitées régionalement ou nationalement.

Des précisions sont apportées notamment pour le développement des outils informatiques : développement de standards, d'interfaces permettant une meilleure dissémination, un meilleur

échange des connaissances et un accès aux données le plus large possible pour les individuels comme pour les organismes.

**Projets précis :** une quinzaine d'intentions de réalisations concrètes sont listées dans le mémorandum, sans précision de groupes d'organismes auxquels ils s'appliquent (la botanique n'est pas explicitement citée par exemple). Citons parmi ces intentions « l'appui du développement d'un catalogue électronique des noms de tous les organismes connus », « le développement d'une bibliothèque électronique sur les données touchant la biodiversité », « le partage des ressources informatiques permettant le traitement et le stockage d'un très grand volume de données » etc...

**Membres :** tout pays et organisme gouvernemental ou non-gouvernemental peut signer la mémorandum du *GBIF*. Deux statuts sont autorisés, en fonction de l'engagement financier des participants : le statut de participant avec droit de vote (Voting participant), sous réserve de contribution financière et le statut de membre associé, sans droit de vote (Associates participant). Trois représentants français ont le statut de « Voting participant »).

**Moyens :** financements des membres votants, lien avec le *GEF* .

## Species 2000

<http://www.sp2000.org/>

**Origine :** programme créé à l'initiative de *L'IUBS* (international Union of Biological Sciences), du *CODATA* (committee on data for Science and Technology) et de *IUMS* (International Union of Microbiological Societies) en Septembre 1994. Puis *Species2000* est approuvé par *l'UNEP* (United Nations Environmental Programs) en 1996-97, qui rappelle le est à l'origine de la préparation de la conférence de Rio en 1992. Ce programme est associé depuis à la *CBD* et au *Clearing House Mechanism*.

**Membres et mandataires :** *Species2000* est une sorte de fédération d'organismes développant des bases de données taxonomiques. Pour la botanique, des représentants des bases de données sur les Palmiers, les Légumineuses (Fabaceae, Cesalpinoaceae, Mimosaceae), les Cactaceae, les Ombellifères et les plantes fossiles ont participé à la réunion de création de ce programme en 1996.

**Objectifs et réalisations :** le programme *Species2000* veut fournir un index valide pour toutes les espèces connues afin de disposer d'une référence électronique stable pour tous les projets d'inventaires, permettant donc leur comparaison ... Un portail vers tous les sites traitant de chacun des groupes taxonomiques est également proposé.

Chaque année la liste *Species2000* est révisée et réactualisée. Elle est disponible sur Internet ou sur CD-ROM.

**Moyens financiers :** les bases de données *Species2000* et le projet *Species2000* lui même sont subventionnés par le *GEF* et *L'UNEP*.

**Liens avec d'autres organismes ou programmes :** *Species2000* est évidemment lié à tous les organismes autour de la *CBD* (*UNEP*, *GEF*), au *GBIF* ainsi qu'aux autres projets ou réalisations de bases de données (*RefTax*, *EMP*, *SPP*, *GPL*...).

## SPP : The species plantarum project

<http://iopi.csu.edu.au/iopi>

**Origine** : projet soutenu par l'IOPI.

**Objectif** : proposer un ensemble de données taxonomiques minimales pour chaque espèce de plantes au niveau mondial : traitement de la nomenclature (synonymie, auteurs, publications), de descriptions succinctes, de la distributions etc...

**Moyens** : les Australiens sont très activement impliqués dans ce projet, relié directement aux ouvrages papier « Flora of the World » et au glossaire publié pour « Flora Australia ».

## GPL : the Global Plant check-list

<http://iopi.csu.edu.au/iopi>

**Origine** : principal projet soutenu par l'IOPI

**Objectif** : produire une check-list complète des 300 000 taxons de plantes vasculaires, représentant environ 1 000 000 de noms nomenclaturaux.

**Moyens** : un comité GPCP organise le travail autour de ce projet. Un modèle de structuration des données a été développé et est proposé. (<http://www.bgbm.fu-berlin.de/iopi/iopimodel73/7301root.htm>).

**Liens avec d'autres projets** : directement lié au projet *Species Plantarum*, ainsi qu'à *Species2000*.

## EMP (ou Euro+Med) project for Euro-Mediterranean Plant Diversity

<http://www.euromed.org.uk>

**Origine et mandataires** : le projet EMP a été présenté en juin 1998 par un consortium de botanistes internationaux auprès de l'Union Européenne, animé par Stephen Jury de l'université de Reading . Il est officiellement approuvé et subventionné à partir du 1er septembre 2000. D'autres organismes non gouvernementaux, comme la Société linnéenne de Londres, la Société Botanique des îles Britanniques, l'atlas FLORA EUROPEA ([http://www.fmnh.helsinki.fi/map/afe/E\\_afe.htm](http://www.fmnh.helsinki.fi/map/afe/E_afe.htm)) le soutiennent financièrement. La France a un représentant dans ce projet et les moyens financiers ont été attribués aux seuls pays suivants : Royaume Uni, Espagne, Allemagne, Italie, Suisse, Finlande, République Tchèque, Grèce, Bulgarie.

**Objectifs** : l'objectif de ce projet est de proposer **une base de donnée en ligne sur les plantes vasculaires d'Europe et de Méditerranée** (pays ayant une façade donnant sur cette mer), comprenant un maximum de renseignements biologiques, écologiques, de répartition et taxonomiques. Les problèmes de nomenclature qui ont émergé après la publication en trois volumes de la « *MedChecklist* », liste des taxons méditerranéens (Greuter et al., ?), après comparaison avec les noms publiés dans FLORA EUROPAEA, ont poussé les initiateurs de ce projet à étendre l'aire d'étude des plantes concernées. Ainsi l'Europe (incluant la Macaronésie et le Caucase) et la Méditerranée sont englobées dans EMP afin de résoudre ces problèmes nomenclaturaux.



**Moyens :** *EMP* s'organise autour d'un noyau taxonomique (« taxonomic core ») fourni par les taxa de l'atlas de FLORA EUROPAEA, autour duquel se greffent :

- des bases satellites (« satellite databases ») plus détaillées (notamment la base de données d'information génétique OPTIMA sur les chromosomes).
- une base de données résumées (« summary beads ») également reliée au noyau taxonomique qui fournira des résumés de données vérifiées par des spécialistes, sur divers thèmes (atlas de répartition, données caryologiques etc...).

Les informations taxonomiques sont périodiquement réévaluées et remises à jour. La mise en place de ce programme s'articule autour de 9 thèmes de travail (« Workpackage ») animés chacun par un coordinateur d'un des pays concernés.

Lien avec les autres projets et groupements : *EMP* est directement en relation avec *Species2000* et le projet *Medcheck list* a été initié par *OPTIMA*.

## **ENHSIN : European Natural History Specimen Information Network**

<http://www.nhm.ac.uk/science/rco/enhsin/index.html>

**Origine :** *ENHSIN* est un projet initié à travers le programme de la Communauté Européenne « Improving the Human Research Potential and the Socio-economic Base » (<http://www.cordis.lu/improving/>).

**Objectifs :** *ENHSIN* se fixe comme objectifs de proposer et développer des programmes de bases de données communes et interactives pour la gestion des informations sur **tous les spécimens** disponibles dans les collections européennes. Le premier meeting a eu lieu en janvier 2000 à Londres, initié par le Muséum d'histoire naturel de Londres (NHM).

**Membres :** on retrouve de grandes institutions , le *MNHN* (Paris), le jardin botanique de Berlin-Darlhém (*BGBM*), le Muséum d'histoire naturelle de Madrid, le musée zoologique de Copenhague, le Royal Botanical Garden de Kew (*RBGK*).

**Liens :** *ENHSIN* a un lien naturel avec *Biocise*, bien que différant légèrement quant au niveau d'appréhension (voir ci-dessous), ainsi qu'avec de nombreux partenaires de *Species2000*.

## **Biocise : Ressource Identification for a Biological Collection Information Service in Europe**

<http://www.bgbm.fu-berlin.de/biocise/>

**Origine :** *BioCise* est également un programme émanant de la Communauté Européenne mais à travers la Commission européenne, sa mission recherche (« Research Directorate-General's Mission ») [http://europa.eu.int/comm/dgs/research/index\\_en.html](http://europa.eu.int/comm/dgs/research/index_en.html) et le « Biotechnology Programme (*BIOTECH*) <http://europa.eu.int/comm/research/biot1.html> mis en place dans le cadre du « Fourth Framework Programme (1994-1998) ». <http://europa.eu.int/comm/research/fp4.html> , plan quadriennal encadrant et subventionnant tous les travaux de recherche et de développements

technologiques initiés en Europe. Remarque : actuellement nous sommes dans le cadre du Fifth Framework Programme (1998-2002) et au projet *BioCise* succède maintenant le projet *BioCASE* (Voir ci-dessous).

**Objectifs :** Les informations contenues dans la multitude des collections détenues et gérées par différentes structures (muséums, herbiers, universités, associations, autres organismes...). sont très éparpillées : chaque collection renferme des informations uniques, portant par exemple sur la localisation, l'écologie ou sur la biologie des espèces dont les spécimens sont stockés. Jusqu'à présent, il n'y avait que peu de communications et d'échanges entre les institutions et seuls des déplacements de scientifiques au sein même des collections ou des échanges de spécimens permettaient un travail transversal et synthétique. *Biocise* souhaite proposer un **système commun d'accès à toutes les informations contenues dans les collections européennes.**

**Membres :** des scientifiques de diverses universités européennes participent au comité d'action concertée de *Biocise*.

**Moyens :** plus de 2800 organismes ont été sollicités par questionnaire, 483 ont répondu favorablement, et 40% des données déjà informatisées concernent les herbiers. Chacun fonctionne souvent sous son propre système de gestion de base de données et l'un des défis majeurs est de développer un outil informatique permettant l'interrogation de tous ces systèmes et leur parfaite communication.

Les réflexions émanant du projet *CDEFD* (a Common Datastructure for European Floristic Databases) <http://www.bgbm.fu-berlin.de/CDEFD/CollectionModel/cdefd.htm> ont servi d'évaluation de départ des problèmes soulevés par une informatisation commune des collections. Un modèle de représentation des données a été proposé, et finalement publié par *Biocise*. La complexité de ces données fragmentaires mais paradoxalement la possibilité réelle de leur représentation par un système commun a été mise en évidence par ce premier travail (Berendsohn et al., 1999).

**Liens avec les autres programmes :** La différence entre *Biocise* et *ENHSIN* se situe dans l'éventail des champs couverts (*ENHSIN* couvre également la minéralogie) ainsi que dans l'échelle de réflexion, *Biocise* se situant à l'échelle des collections, *ENHSIN* plus précisément à celle du spécimen. *BIOCISE* est maintenant remplacé par *BioCASE*.

## **BioCASE a biodiversity collection access service for Europe**

<http://www.biocase.org>

**Origine :** *BioCASE* succède donc à *Biocise*, dans le cadre du Cinquième plan cadre Européen (Fifth Framework Programme, voir ci-dessus *Biocise*). Ce projet a démarré le 1er novembre 2001.

**Objectifs :** les objectifs de *BioCASE* reprennent ceux de *BioCise* et de *Enhsin*, ils sont élargis et visent à mettre en connexion toutes les informations contenues dans les collections européennes, à tous les niveaux. La notion de propriété intellectuelle des données est abordée, et il est explicitement stipulé que celles-ci resteront créditées aux organismes qui les détiennent.

**Moyens :** *BioCASE* tire les enseignements de *BioCise* et *ENHSIN*. On passe à l'échelle des « Metadonnées », sorte de base de données compilation de base de données. La connexion de ces

bases hétérogènes pose des problèmes techniques et informatiques : rendre les informations homogènes sans modifier les structures des bases originelles, réduire les temps d'accès entre les divers serveurs et les consultants (clients) etc...Mais une des difficultés majeure à laquelle se heurtent les personnes travaillant sur ce type de projet est la réticence de nombreux organismes ou institutions à partager leurs données, même s'il s'agit de données sur les collections et non pas de données d'inventaires.

## Projets nationaux

### Reftax

<http://www.mnhn.fr/base/reftax>

**Origine** : le projet *Reftax* initié par le *MNHN*, Muséum National d'Histoire Naturel de Paris ( <http://www.mnhn.fr> ) a débuté en 1999.

**Objectifs et réalisations** : *RefTax* veut fournir en libre accès pour tous, institutions comme individuels, un référentiel taxonomique informatisé pour tous les organismes vivants, en priorité ceux observés sur le territoire français et métropolitain. Dans ce référentiel, les noms vernaculaires, lorsqu'ils sont disponibles, seront reliés aux noms scientifiques. Les références bibliographiques de tous les noms cités seront également fournies.

Cette base sera bien sûr disponible sur Internet et interrogeables selon divers modes.

**Mandataires** : le projet *Reftax* est organisé autour de trois comités : le comité de pilotage, le comité éditorial, le comité scientifique. Seul le premier est actuellement constitué : une quinzaine de personnes liés aux laboratoires du *MNHN*, à l'Institut Française de Systématique (*IFS*) hébergée au *MNHN*, à l'université Pierre et Marie Curie (*UPMC*) sont chargées d'assurer le démarrage de ce projet et notamment de l'évaluation des dossiers présentés.

**Moyens** : le projet *RefTax* s'appuie sur tous les acteurs institutionnels ou privés travaillant autour de la diversité. Un appel d'offre a été lancé pour la fourniture de listes de noms répondant aux critères de *Reftax* et est évalué tous les six mois par le comité de pilotage.

## Conclusion

Tous ces projets proposent déjà des données consultables en ligne, mais dont les données sont loin d'être exhaustives...mais par définition, ces bases de données doivent continuer à s'enrichir au fil du temps.

Quelles idées retenir de cette profusion de projets, dont certains semblent se recouvrir ?

L'application de la *CBD* a ouvert un vaste champ de recherche à l'interface biologie-informatique, répondant à de réels besoins de la communauté scientifique au sens large (scientifiques institutionnels et naturalistes non institutionnels - le terme « amateur » est souvent mal approprié...). La hiérarchisation et l'imbrication des structures démissionnaires et gouvernementales (Nations unies, Europe, gouvernements nationaux, structures décisionnelles régionales) dont dépendent le

fonctionnement des différentes institutions et associations privées, si ce n'est totalement (muséums...) tout au moins pour une part de leur financement (rares sont les associations naturalistes non subventionnées d'une manière ou d'une autre par une structure institutionnelle) explique l'émergence simultanée de tous ces projets. Au delà de la complexité de cette arborescence, il faut plutôt retenir la richesse des données potentiellement accessibles et la volonté de les mettre en commun. De la multitude de propositions de systèmes de gestion des données émergera sans doute à l'usage une idée plus précise de leur structuration optimale et des outils informatiques les mieux adaptés.

Cependant tout n'est pas si rose dans le royaume de la BD...et de nombreux écueils restent à éviter et à dépasser. Citons, d'après Nicolas Bailly (Bailly, 2001)

- Les problèmes **liés à la propriété intellectuelle** des données : tout le monde est - il prêt à « jouer le jeu » (voir remarques sur le programme *BioCASE*)? Certains ou certaines institutions préfèrent exploiter les informations , sans communiquer les données brutes. Les raisons en sont diverses, les naturalistes étant souvent échaudés par une utilisation de leurs prospections sans la moindre citation des sources, reconnaissance minimale du travail effectué.
- Le problèmes liés à la **qualité des données** : faut-il croire tout ce qui est écrit ? des erreurs d'identification subsistent dans les herbiers, des données nomenclaturales sans références ou avec des références erronées sont inexploitable etc...qui vérifie ?
- Le problème de la **diffusion des données**
- Le problème de la **maintenance des données** : qui mettra à jour, corrigera etc...

Plusieurs axes de recherche vont dans ce sens, l'un de ceux développés actuellement en bio-informatique est celui de la « saisie coopérative sur le Web » (BOSSY R., thèse en cours (Bossy, 2001) <http://lis.snv.jussieu.fr/~bossy/>), qui permet l'actualisation contrôlée des bases de données. Mais là encore la collaboration continue des systématiciens, des détenteurs de données et des informaticiens est plus que jamais nécessaire.

**Catherine REEB**

Laboratoire Informatique et Systématique, UPMC, 12 rue Cuvier, 75005 Paris

## Bibliographie

- Bailly, N. (2001) In *Journées nationales de la SFS*, Académie des sciences.
- Bailly, N (2001) Compte rendu de la journée SFS du 27/02/01, *Bulletin de la SFS*, **26**, pp 5-10
- Berendsohn, W. G., Anagnostopoulos, A., Hagedorn, G., akupovic, J., Nimis, P. L., Valdés, B., Güntsch, A., Pankhurst, R. J. and White, R. J. (1999) *Taxon*, **48**, 511-562.
- Bisby F.A., Russell G.F. and Pankhurst R.J. (eds), 1993. *Designs for a global plant species information system*. Oxford, Clarendon Press. 350 p. (Systematics Association, Special volume No. **48**).
- Bossy, R. (2001) In *Journées Ouvertes Biologie Informatique et Mathématiques (JOBIM2001)*(Ed, Laurent Duret, C. C., and Thomas Schiex,) Toulouse, pp. 43-51.

- Brummitt, R. K. and Powell, C. E. (1992) *Authors of plant names*, TDWG standard, Royal Botanical gardens Kew.
- Conn, B. (1996) *Herbarium Information Standards and Protocols for interchange of data Version 3*, Royal Botanical garden, Sydney.
- Delobel, C., L  cluse, C. and Richard, P. (1991) *Bases de donn  es: des syst  mes relationnels aux syst  mes    objets*, Inter  ditions.
- Gardarin, G. (1999) *Bases de donn  es*, Eyrolles, paris.
- Greuter, W., H., B. and Long, D. (?) *Medcheklist*, .
- Lebbe, J. (1996) *Informatique et syst  matique*, SFS, P. d. l. BIOSYSTEMA, , paris.
- Pankhurst, R. J. (1978) *Biological identification. The principle and practice of identification methods in biology.*, Edward Arnold., London.
- Stajano, F. (1998) , Vol. 2001 Olivetti and oracle research laboratory, .

## Applications

- DELTA(1980) <http://biodiversity.uno.edu/delta/>
- X-PER (1983) <http://lis.snv.jussieu.fr/Xper/>.

## Organismes

- UNEP (United Nations Environment Programme <http://www.unep-wcmc.org> )
- GEF (Global Environment Facility) <http://www.undp.org/gef>
- l'UNDP (United Nations Development programme) <http://www.undp.org>
- la World Bank <http://www.worldbank.org>
- TDWG : Taxa Database Working Group <http://www.tdwg.org>
- GTI : Global Taxonomy Initiative <http://www.biodiv.org/programmes/cross-cutting/taxonomy>
- OECD ou OCDE en fran  ais (organisation for economic cooperation and development <http://www.oecd.org>)
- Commission europ  enne, sa mission recherche (« Research Directorate-General's Mission ») [http://europa.eu.int/comm/dgs/research/index\\_en.html](http://europa.eu.int/comm/dgs/research/index_en.html)
- « Biotechnology Programme (BIOTECH) » <http://europa.eu.int/comm/research/biot1.html>
- « Fourth Framework Programme (1994-1998) ».  
<http://europa.eu.int/comm/research/fp4.html>
- OPTIMA : <http://www.bgbm.fu-Berlin.de/OPTIMA/ITN/recommendations.htm>
- IOPI : <http://iopi.csu.edu.au/iopi/>
- IAPT : International Association for Plant taxonomy <http://www.bgbm.fu-berlin.de/iapt/default.htm>
- CoDATA : Committee on Data for Science and Technology  
<http://www.codata.org/welcome.html>
- SFS : Soci  t   Fran  aise de syst  matique <http://sfs.snv.jussieu.fr/>

## Programmes

- Convention sur la Biodiversité (*CBD*) (<http://www.biodiv.org>)
- Le *GBIF* Global Biodiversity Information Facility <http://www.gbif.org>
- *ENHSIN* : European Natural History Specimen Information Network  
<http://www.nhm.ac.uk/science/rco/enhsin/index.html>
- programme de la Communauté Européenne « Improving the Human Research Potential and the Socio-economic Base » (<http://www.cordis.lu/improving/>).
- *BIOCISE* : Ressource Identification for a Biological Collection Information Service in Europe <http://www.bgbm.fu-berlin.de/biocise/>
- *BioCASE* : <http://www.biocase.org>
- *Species 2000* : <http://www.sp2000.org/>
- *Reftax* <http://www.mnhn.fr/base/reftax/pub/index2.html>
- *FLORA EUROPEA* ([http://www.fmnh.helsinki.fi/map/afe/E\\_afe.htm](http://www.fmnh.helsinki.fi/map/afe/E_afe.htm))
- *GPCL* : The global plant check-list : <http://iopi.csu.edu.au/iopi/iopigpc1.html>
- *SPP* : The species plantarum project <http://iopi.csu.edu.au/iopi/iopispp1.html>
- *EMP* (ou Euro+Med) project for Euro-Mediterranean Plant Diversity  
<http://www.euromed.org.uk>
- *CDEFD* (a Common Datastructure for European Floristic Databases) <http://www.bgbm.fu-berlin.de/CDEFD/CollectionModel/cdefd.htm>