

## **Réseau Bases de Données**

**Présenté auprès de la Mission aux Ressources et Compétences Techniques du  
CNRS**

**2012**

## Sommaire

Pourquoi créer un réseau national ? .....	3
Quelles sont les personnes concernées ? .....	5
Quels sont les objectifs du réseau ?.....	6
Le comité de pilotage national (CPN).....	7
Rôle .....	7
Fonctionnement .....	7
Composition .....	7
Responsable/Animateur.....	7
Membres du comité de pilotage au démarrage du réseau.....	7
Les groupes de travail.....	8
Actions à réaliser .....	11
Budget prévisionnel 2012-2013 .....	11
Les partenaires institutionnels .....	11
Annexes.....	12

## Pourquoi créer un réseau national ?

L'idée de création d'un réseau dédié aux questionnements autour des bases de données est née suite aux résultats de deux enquêtes :

- Une 1ère enquête s'adressant à l'ensemble des instituts du CNRS réalisée par le Ministère de l'Enseignement supérieur et de la Recherche (Françoise Thibault) en 2009 sur une « Etude de cas sur les systèmes de données numériques de recherche » et à laquelle 282 unités ont répondu et 690 bases ont été déclarées (Doc 1).
- Une 2ème enquête réalisée au sein de l'INEE-CNRS en 2010 (Stéphanie Thiébault et Yvan Lagadeuc) a permis de préciser les problèmes rencontrés par les personnels travaillant sur les bases de données.

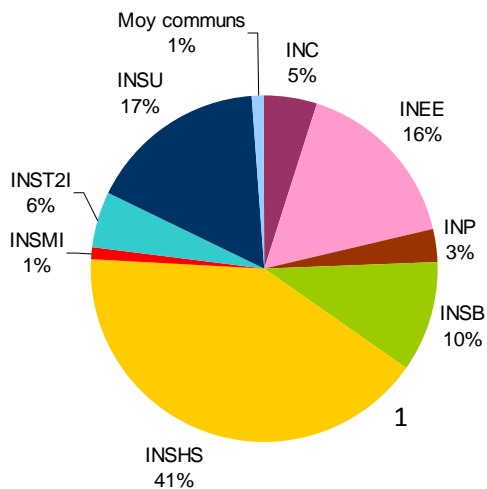
Ces différentes enquêtes ont permis de mettre en évidence des questionnements et des problématiques communes à toutes les disciplines :

- Des problèmes de pérennisation à la fois institutionnelle, financière, structurelle
  - convention et hébergement
  - insuffisance de moyens financiers, projet sur des durées limitées et ne permettant pas le maintien de certaines bases
  - problèmes de continuité des services avec l'utilisation de contrats non pérennes
- Un manque de communication et un isolement important des personnels affectés à la création ou la gestion des bases de données.
- Une dispersion des ressources humaines et financières
- Une grande diversité des interlocuteurs.  
Les questionnements liés aux bases de données concernent les ingénieurs « Bases de données » comme les **chercheurs qui participent activement à la mise en place de dictionnaires de données (métadonnées et thésaurus) et qui jouent également un rôle d'ingénierie en base de données.**
- Des problèmes d'interopérabilité des bases de données.  
La multiplication des acteurs intervenant lors de la création d'une base de données (Ingénieurs informaticiens, Bio informaticiens, Chercheurs, Enseignants), a pour effet **une grande confusion** entre la structuration de la base (relationnelle ou objet), le choix des métadonnées/thésaurus et les données contenues dans la base à la fois dans la **construction** et dans la **protection juridique**.
- **Une inquiétude croissante concernant l'adaptation au droit français de la directive Européenne INSPIRE.** La mise en place en 2007 de cette directive rend obligatoire à tous les états membres de l'UE, la mise à disposition de toutes les données publiques concernant des séries de données géographiques.
- Des besoins importants d'informations dans le domaine juridique concernant la protection des données et des bases de données.
- Une meilleure valorisation du travail scientifique lié à la création d'une base de données et une reconnaissance de ce travail comme production scientifique (en particulier en ce qui concerne les rapports d'activité).

Les résultats de cette enquête s'accompagnent d'une demande croissante des unités CNRS d'aide pour la réalisation de bases de données. Pour répondre à ces besoins de plus en plus importants, quelques centres informatiques dédiés à la gestion de données ont été créés, mais, dans la majorité des cas, la

réponse a été d'engager ponctuellement un personnel ITA bases de données au sein d'unité de recherche. Ces recrutements s'accompagnent pour les agents d'une grande diversité de tâches à effectuer, allant du simple dépannage informatique à la création de bases de données complexes nécessitant des formations complémentaires.

Ces constatations émergent au sein de tous les instituts du CNRS et quelque soit les disciplines. Il est donc nécessaire de mettre en place un réseau de questionnement commun à toutes les équipes du CNRS et aux autres établissements avec la création d'un réseau « bases de données » (à la fois thématique et de métiers).



Document 1 : Répartition par institut des unités ayant répondu au questionnaire de l'enquête réalisée en 2009 (1/Données pondérées 2/Données brutes)

Instituts	Nombre d'unités ayant répondues
Institut de chimie	35
Institut écologie et environnement	112
INP	22
Institut des sciences biologiques	71
Institut des sciences humaines et sociales	284
Institut national de mathématiques et de leurs interactions	6
INST2I	38
Institut national des sciences de l'univers	114
Moyens communs	8

## Quelles sont les personnes concernées ?

Les intervenants du réseau peuvent provenir de tous les établissements de la recherche publique et de tous les corps d'activités professionnelles aussi bien chercheurs, enseignants-chercheurs, ingénieurs ou techniciens. Toutes les disciplines sont concernées, même si les communautés au démarrage du réseau sont pour le moment principalement axées sur les domaines environnemental et sciences humaines et de la société. Au vue de l'enquête de 2009 (cf. **doc.1**), on peut évoquer l'INSU, l'INS2I, l'INSB, L'INC, l'INP, l'INSMI...

La communauté spécialisée dans les bases de données se compose de personnels provenant de divers établissements et institutions : CNRS, universités (Pôle de recherche et d'enseignement supérieur ou PRES, Observatoire des sciences de l'univers ou OSU), établissements dépendant du Ministère de l'écologie, du développement durable, des transports et du logement et du Ministère de l'Enseignement Supérieur et de la Recherche (Muséum national d'Histoire naturelle, INRA, INSERM).

La création d'une base de données nécessite l'intervention de différents types de personnels en fonction du niveau d'intervention dans le processus de réalisation :

### Partie REFLEXION/GENESE

Cette étape consiste à identifier les besoins et les problèmes des utilisateurs de la future base de données. Il doit permettre d'aboutir aux étapes suivantes : définition des dictionnaires de données (partie descriptive) et schéma de la base de données (partie structurelle).

### Partie DESCRIPTIVE

Choix et description des métadonnées : Choix des *Métadonnées « Primaires »* déjà préexistantes et définies par différents organismes d'état (ex : listes taxonomiques, données de géo-localisation, données administratives) et des *Métadonnées « secondaires »* c.à.d. propres à chaque recherche et qui nécessitent une description approfondie. Définition du dictionnaire de données (thésaurus, mots clés).

### Partie STRUCTURELLE

Structuration technique : squelette de la base de données (ex : choix du SGBD, langage de programmation, réalisation du portail d'accès aux données) et choix de l'interface de consultation de la base.

### Partie FONCTIONNEMENT

Consultation, utilisation, sauvegarde et saisie des données : informations sur les droits des données contenues et sur la base elle-même, sur le mode de consultation, de stockage et de saisie de ces données.

### Partie PERENNISATION

Pérennisation des bases de données : mise à jour des systèmes de gestion des bases de données, maintien de l'activité des bases et archivage lorsqu'elles sont inactives.

## **Quels sont les objectifs du réseau ?**

1- Fédérer les agents dont les missions sont liées aux bases de données en mettant en place un réseau de communication, d'échanges et de projets permettant d'apporter une aide et une dynamique dans tous les domaines techniques nécessaires à la conception, à la réalisation, à l'exploitation ou à l'adaptation optimale des bases de données : veille technologique, entraide technique, prospectives techniques, offres d'emplois...

2- Mettre en place des groupes de réflexions sur les dictionnaires de données (jeux de métadonnées et thésaurus) qui peuvent être utilisés pour la réalisation des bases de données et les mettre à la disposition de la communauté.

3- Harmoniser des actions avec les animateurs des grands projets scientifiques multidisciplinaires dans lesquels une ou plusieurs bases de données jouent un rôle fondamental. Essaimer l'expertise du réseau et interagir de façon ouverte avec toute entité active dans le domaine des bases de données.

4- Connaître les forces et les compétences des personnels (statutaires ou contractuels) en réalisant un état des lieux auprès des unités CNRS et proposer un annuaire des compétences au sein du réseau. Cet annuaire de compétences sera ouvert à d'autres établissements après accord des partenaires institutionnel.

5- Réaliser des fiches-conseils pour aider à la création d'une base de données et permettre son interopérabilité avec les bases d'une même discipline.

6- Proposer et organiser des formations permettant à la communauté de maîtriser les meilleures innovations techniques adaptées au contexte de la recherche.

7- Proposer des moyens de publications pour favoriser la visibilité du réseau et produire des recommandations

8- Proposer à tous les intervenants des bases de données (ITA Bases de données, Chercheurs et Enseignants) un guide des bonnes pratiques des bases de données, validé par l'ensemble de la communauté.

9- Favoriser la cohésion et les échanges au sein du groupe en proposant périodiquement une réunion regroupant l'ensemble des membres du réseau...

## **Le comité de pilotage national (CPN)**

### **Rôle**

Les membres du comité de pilotage ont pour rôle :

- d'animer le réseau
- d'identifier les attentes des membres du réseau et des communautés scientifiques et techniques concernées
- de diffuser des éléments de prospectives
- de dynamiser l'implication de la communauté dans des appels à projets technologiques
- de vérifier l'impact et la visibilité du réseau au sein de nos organismes et en particulier dans les laboratoires
- de veiller à la communication interne et externe du réseau
- de représenter le réseau
- de préparer le budget prévisionnel du réseau

### **Fonctionnement**

Le comité de pilotage se réunit environ deux fois par an sous la direction de son responsable. Le mode distant (ou visioconférence) sera développé afin de limiter les déplacements au strict nécessaire tout en favorisant des échanges réguliers au sein du groupe. Chaque réunion fait l'objet d'un compte rendu ou relevé de décisions. Un rapport d'activité est réalisé tous les deux ans.

Afin de conserver des liens forts avec les différents instituts du CNRS ou les différents établissements partenaires, il est possible d'inviter lors de ces réunions des représentants extérieurs au comité de pilotage.

### **Composition**

La composition du comité de pilotage se fait en fonction des nécessités du réseau, des compétences et des motivations de chacun. Le nombre des intervenants est d'environ quinze personnes (avec la possibilité de nommer des suppléants) pour un mandat en principe de quatre ans (mais qui pourra évoluer plus rapidement au cours de premières années en fonction de l'ouverture disciplinaire du réseau), choisis sur la base du volontariat et en tentant d'obtenir une représentation équilibrée des instituts et de la diversité des fonctions liées aux bases de données (cf. page 6). Des représentants d'établissements partenaires ou bien d'autres réseaux du CNRS peuvent être invités aux discussions du comité de pilotage.

### **Responsable/Animateur**

L'animateur/responsable du réseau est nommé par le comité de pilotage pour une durée de 2 ans renouvelable. Un co-responsable est également désigné, celui-ci pouvant remplacer le responsable en cas d'indisponibilité.

### **Membres du comité de pilotage au démarrage du réseau**

BALY Isabelle baly@mnhn.fr UMR7209 Archéozoologie, archéobotanique : sociétés, pratiques et environnements, MNHN, Paris.

BOURGOIN Thierry, bourgoin@mnhn.fr, UMR7205 Origine, structure et évolution de la biodiversité, MNHN, Paris.

CHARVY Jean-Christophe, j-christophe.charvy@univ-bpclermont.fr, UMR6023 Microorganismes : génome et environnement (MGE).

FANTINO Guillaume, guillaume.fantino@ens-lyon.fr, UMR5600 Environnement, ville, société - Zone Atelier Bassin du Rhône.

FLOQUET Thomas, thomas.floquet@univ-bpclermont.fr, UMR6042 Laboratoire de Géographie physique et environnementale (GEOLAB).

GARNIER Eric, eric.garnier@cefe.cnrs.fr, UMR5175 Centre d'Ecologie Fonctionnelle et Evolutive (CEFE).

LE MOAL Françoise, francoise.le-moal@univ-rennes1.fr, UMR6553 Ecosystèmes, biodiversité, évolution (ECOBIO), OSU Renne, Renne.

MAIGNAN Maxime, maignan@eccorev.fr, FR3098 Ecosystèmes continentaux et risques environnementaux – (ECCOREV).

MARTIN Chloé, chloe.martin@mnhn.fr, UMS3468 Bases de données sur la Biodiversité, Ecologie, Environnement et Sociétés (BBEES), MNHN, Paris.

MICHEL Kristell, kristell.michel@ens-lyon.fr, UMR5600 Environnement, ville, société.

PENNEC Flora, pennec@mnhn.fr, UMR7206 Eco-Anthropologie et ethnobiologie.

PIGNON-MUSSAUD Cécilia, cecilia.pignon-mussaud@univ-lr.fr, UMR6250 Littoral, Environnement et Sociétés (LIENSs).

PRESSAC Jean-Baptiste, jean-baptiste.pressac@univ-brest.fr, Centre de Recherche Bretonne et Celtique, Université de Bretagne Occidentale, Brest.

ROTEREAU Armand, rotereau@eccorev.fr, FR3098 Ecosystèmes continentaux et risques environnementaux (ECCOREV).

SILVESTRE Marie, marie.silvestre@upmc.fr, FR3020 Fédération Ile-de-France de Recherche sur l'Environnement (FIRE).

## **Les groupes de travail**

Les groupes de travail les plus avancés pour le moment concernent essentiellement les thésaurus et dictionnaires de données et plus particulièrement la définition de la directive européenne Inspire et son application sur les données de la recherche. Des contacts ont déjà été pris avec le ministère de l'écologie, du développement durable, des transports et du logement (MEDDL) afin d'organiser dès la fin de l'année une présentation d'Inspire et une formation début 2013.

Les groupes de travaux « dictionnaires de données », « évolutions technologiques, veille documentaire et bonnes pratiques » et « Informations juridiques » répondent également à une demande forte de la communauté et sont des thématiques qui concerneront tous les instituts du CNRS. Plusieurs membres du comité du pilotage se sont portés volontaires pour gérer ces groupes et des premiers résultats pourront être visible dans le courant de l'année 2013.



**Groupe de travail « Métadonnées pour l'information géographique, INSPIRE » :**

- Normes : définition des normes ISO existantes (ISO 19115, 19119, 19110)
- Explication de la Directive INSPIRE
- Publication d'une documentation et de recommandations

*Responsables proposés : Cécile Maréchal et Chloé Martin*

**Groupe de travail « Dictionnaires de données » :**

- Recenser les jeux de métadonnées existants (autres que géographiques), proposer de nouveaux jeux de métadonnées à la communauté.
- Proposer une documentation expliquant comment décrire ses données selon les normes existantes (Dublin Core ou INSPIRE adaptée) (*Responsable proposé : Chloé Martin et Cécile Maréchal*)
- Publier en ligne de la documentation et des jeux de métadonnées (ou les liens vers les institutions qui les proposent).
- Définir des thésaurus par discipline et les proposer au téléchargement.
- Créer une charte d'utilisation de ces Métadonnées et des thésaurus (charte de bonne conduite).
- Inciter la direction du CNRS à démarrer dès maintenant une réflexion sur la politique de diffusion des données en vue d'une mise en conformité de l'établissement face à la directive INSPIRE.
- Réflexion sur la taxonomie (*Responsable proposé : Thierry Bourgoïn*)

**Groupe de travail évolutions technologiques, veille documentaire et bonnes pratiques :**

- Réflexion sur l'évolution des techniques (bases orientées objet, NoSQL, framework)
- Mise en place d'une documentation/publication mise à la disposition sur le net.
- Outils existants
- Solutions de stockage des bases de données

*Responsables proposés : Jean-Baptiste Barreau et Jean-Baptiste Pressac*

**Groupe de travail « Informations juridiques » :**

Quelles législations ? Protection de la structure des bases, protection des données, droits logiciels (droit d'auteurs, droit sui generis).

Publication de documentations et de recommandations

**Groupe de travail besoins en formations logiciels/techniques :**

Le réseau permet d'offrir à ses membres la possibilité de se former à des nouvelles technologies en

lien avec les bases de données en proposant des formations sur l'utilisation de logiciels et l'apprentissage de langages.

Pôle « compétences et d'excellence » :

- Créer un réseau de membres acceptant de répondre ponctuellement aux problèmes techniques en proposant un listing des compétences de chacun.
- Proposer aux instituts une expertise professionnelle dans la création et la gestion des bases de données.

Groupe de travail « science participative »

- Réflexion sur la validation des données issues de la science participative
- Comment recueillir les données ?

## Actions à réaliser

- 1- Mise en route progressive des groupes de travail (missions, objectifs et livrables)
- 2- Réalisation du site internet support de la documentation et des recommandations et d'un espace réservé pour le dépôt des offres d'emplois.
- 3- Mise en place de la stratégie de formation pluriannuelle (collecter les demandes et établir un plan de formation pour l'année à venir).
- 4- Création d'une plate-forme d'échanges (listes de discussion, espace de partage de documents techniques, espace pour stocker les comptes-rendus des réunions et les présentations, espace de publication de documents).
- 5- Mettre en place des actions de communication auprès des unités des différents instituts du CNRS afin de diffuser les travaux du réseau et de recruter toutes les personnes intéressées par nos problématiques.

### **Actions à lancer dès 2012**

- 1- Organisation d'une réunion en automne 2012 présentant les différents groupes de travail et leurs objectifs pour permettre aux membres du réseau de s'inscrire dans les différents groupes. Cette journée sera également l'occasion de présenter les objectifs de la directive INSPIRE et les outils proposés par le MEDDLT.
- 2- Lancement d'une formation dont l'objectif serait de former des formateurs à la directive INSPIRE qui pourraient ensuite former les personnels CNRS qui le souhaitent.
- 3- Lancement du groupe de travail « INSPIRE » sous la direction de Cécile Maréchal et Chloé Martin.

## Budget prévisionnel 2012-2013

Journée d'étude prévue pour l'automne 2012 (1 jours) : 6000€ (environ une cinquantaine de personnes)

Formation (2 jours) : 1000€ (rémunérations des formateurs pour les journées de formation INSPIRE début 2013) + 1000€ de financement des missions des participants (environ 10 personnes)

Budget missions : 5000€ (pour les réunions du comité de pilotage et des groupes de travail)

## Les partenaires institutionnels

### Partenaires CNRS :

Projet porté par l'UMS3468, Bases de données sur la Biodiversité, Ecologie, Environnement et Sociétés (BBEES)

Promouvoir les Logiciels Utiles Maîtrisés et Economiques dans l'Enseignement Supérieur et la Recherche (PLUME)

Instituts du CNRS : INEE, INSU, INSHS, INIST...

Réseau MRCT :

Medici : Réseau des Métiers de l'Édition Publique scientifique

ResINFO : Fédération des réseaux (métiers) d'Administrateurs Systèmes et Réseaux (ASR) dans le milieu Enseignement / Recherche (Journée Base de données).

Réseau Archéométrie (en cours)

RENATIS : (en cours)

DEVLOG : Réseau de DEVeloppeurs de LOGiciel - Réseau inter-établissements : INRA INRIA Université CNRS (groupe de travail bases de données <http://devlog.cnrs.fr/groupe-de-travail/bases-de-donn%C3%A9es/start>)

Interaction avec le projet MASTODONS (Mission pour l'Interdisciplinarité).

**D'autres réseaux de la MRCT seront contactés via les journées Inter-Réseaux qui auront lieu en novembre 2012.**

Les partenaires extérieurs au CNRS

Les établissements scientifiques dépendant du Ministère de l'Enseignement Supérieur et de la Recherche (MESR) et du Ministère de l'écologie, du développement durable, des transports et du logement (MEDDTL)

Service du Patrimoine naturel ou SPN (MEDDTL, Muséum national d'Histoire naturelle ou MNHN)

Système d'information sur la nature et les paysages ou SINP (MEDDTL)

Fondation pour la Recherche sur la Biodiversité (FRB)

**Annexes**

Liste de laboratoires qui se sont manifestés au démarrage du réseau. Cette liste est ouverte à tous les laboratoires de tous les instituts.

OHM Observatoire Homme-Milieu  
ZA Zone Atelier

FR2424 Station biologique de Roscoff

FR3020 Fédération Ile-de-France de Recherche sur l'Environnement (FIRE)

FR3098 Ecosystèmes continentaux et risques environnementaux - ECCOREV

FRE3350 Laboratoire d'Ecogéochimie des Environnements Benthiques (LECOB)

GDR 2827 Ecologie chimique

GDR2909 Métabolisme de l'arsenic chez les microorganismes (MAMs)

GDR2968 Système d'information phénologique pour l'étude et la gestion des changements climatiques (SIP-GECC)

UMR5119 Ecologie des systèmes marins côtiers (ECOSYM)

UMR5120 Botanique et bioinformatique de l'architecture des plantes (AMAP)

UMR5174 Evolution et diversité biologique (EDB)

UMR5175 Centre d'Ecologie Fonctionnelle et Evolutive (CEFE)

UMR5204 Environnements, Dynamiques et Territoires de la Montagne (EDYTEM)

UMR5244 Ecologie et Evolution des Interactions (2EI)  
UMR5288 Anthropologie Moléculaire et Imagerie de Synthèse  
UMR5290 Maladies Infectieuses et Vecteurs : Ecologie, Génétique, Evolution et Contrôle (MIVEGEC)  
UMR5553 Laboratoire d'écologie alpine (LECA)  
UMR5554 Institut des sciences de l'évolution de Montpellier (ISEM)  
UMR5558 Biométrie et biologie évolutive  
UMR5600 Environnement, ville, société (EVS)  
UMR5608 Travaux et Recherches Archéologiques sur les Cultures, les Espaces et les Sociétés (TRACES)  
UMR6023 Microorganismes : génome et environnement (MGE)  
UMR6042 Laboratoire de Géographie physique et environnementale (GEOLAB)  
UMR6250 Littoral, Environnement et Sociétés - LIENSs  
UMR6282 Biogéosciences  
UMR6553 Ecosystèmes, biodiversité, évolution (ECOBIO)  
UMR6566 Centre de Recherche en Archéologie, Archéosciences, Histoire (CREAAH)  
UMR7138 Systématique, adaptation, évolution SAE  
UMR7156 Génétique moléculaire, génomique et microbiologie (GMGM)  
UMR7179 Mécanismes Adaptatifs : des Organismes aux Communautés  
UMR7204 Conservation des espèces, suivi et restauration des populations  
UMR7205 Origine, Structure et Evolution de la Biodiversité  
UMR7206 Eco-Anthropologie et ethnobiologie  
UMR7207 Paléobiodiversité et paléoenvironnements  
UMR7208 BOREA Biologie des Organismes et Ecosystèmes Aquatiques  
UMR7209 Archéozoologie, archéobotanique : sociétés, pratiques et environnements  
UMR7263 Institut méditerranéen de biodiversité et d'écologie marine et continentale (IMBE)  
UMR7264 Cultures et Environnements. Préhistoire, Antiquité, Moyen Âge (CEPAM)  
UMR7269 Laboratoire méditerranéen de préhistoire Europe-Afrique (LAMPEA)  
UMR7618 Biogéochimie des milieux continentaux BIOEMCO  
UMR7619 Structure et fonctionnement des systèmes hydriques continentaux (SISYPHE)  
UMR8079 Laboratoire d'Ecologie, Systématique et Evolution  
UMR8172 ECOFOG Ecologie des Forêts de Guyane  
UMR8198 Génétique et évolution des populations végétales (GEPV)  
UMR8591 Laboratoire de Géographie Physique  
UMS3194 Centre de recherche en écologie expérimentale et prédictive - Ecotron Ile de France (CEREEP)  
UMS3370 Station Alpine Joseph Fourier  
UMS3462 Systèmes d'observation pour la Conservation des Mammifères et Oiseaux Marins  
UMS3468 Bases de données sur la Biodiversité, Ecologie, Environnement et Sociétés (BBEES)  
UPR1934 Centre d'études biologiques de Chizé (CEBC)  
USR3278 Centre de Recherches Insulaires et Observatoire de l'Environnement de Polynésie Française  
USR3330 Savoirs et Mondes Indiens