

Biomaj

Plate-forme

Bio-informatique

GENOUEST

**Mise à disposition et mutualisation
inter plate-forme des banques de données
biologiques**

Collaboration : INRA Toulouse, INRA Jouy-en-jossas

A.Assi, H.Leroy, O.Filangi

D.Allouche, J-M.Larre

C.Caron, V.Martin

24/10/2006

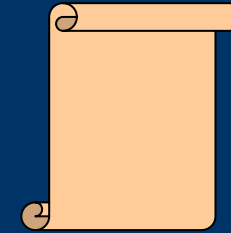
Problématique

- Gestion des banques de données biologiques
 - pré requis indispensable analyse biologique
 - Masse de données volumineuse
 - Fréquence de mise à jours variable
 - Nature et nombre en évolution permanentes
 - Multiples indexations
 - Qualité de service
-
-

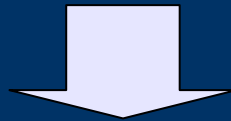
Processus de mise à jour d'une banque biologique

Attributs banques

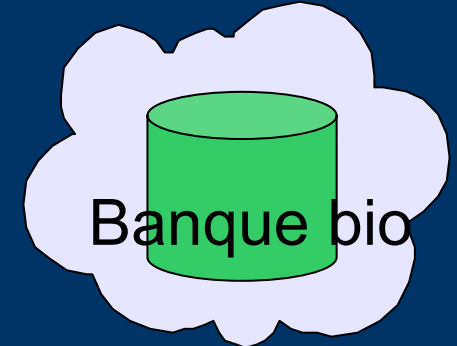
Mirror



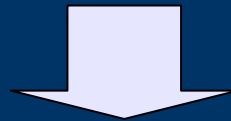
Web



Contrôle



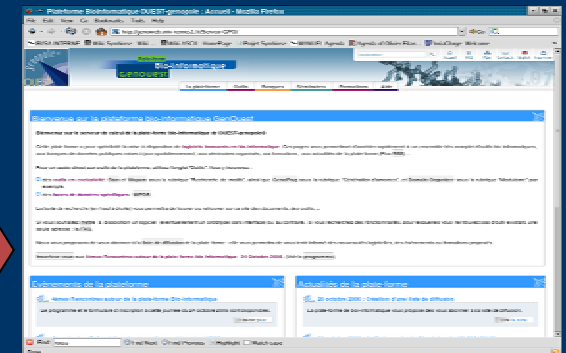
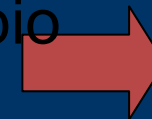
Processus d'indexation



Contrôle

Home

Mise en production



Objectif du projet Biomaj

- Administrer les banques de données biologique via une interface
 - Outil de configuration pour gérer les caractéristiques d'une banque biologique
- Mise en production fiable des données
- Synchronisations régulières des données par rapport au site de référence (ncbi,ebi,...)
 - ➔ Administrateur de banque : Gain de temps
 - ➔ Utilisateur : Qualité du suivi de l'état des données

Origine du projet Biomaj

- **Citrina**

- Communauté Gmod (josh goodman CGB Indiana university)
- Développement supl. Inra toulouse

- **BiomajI**

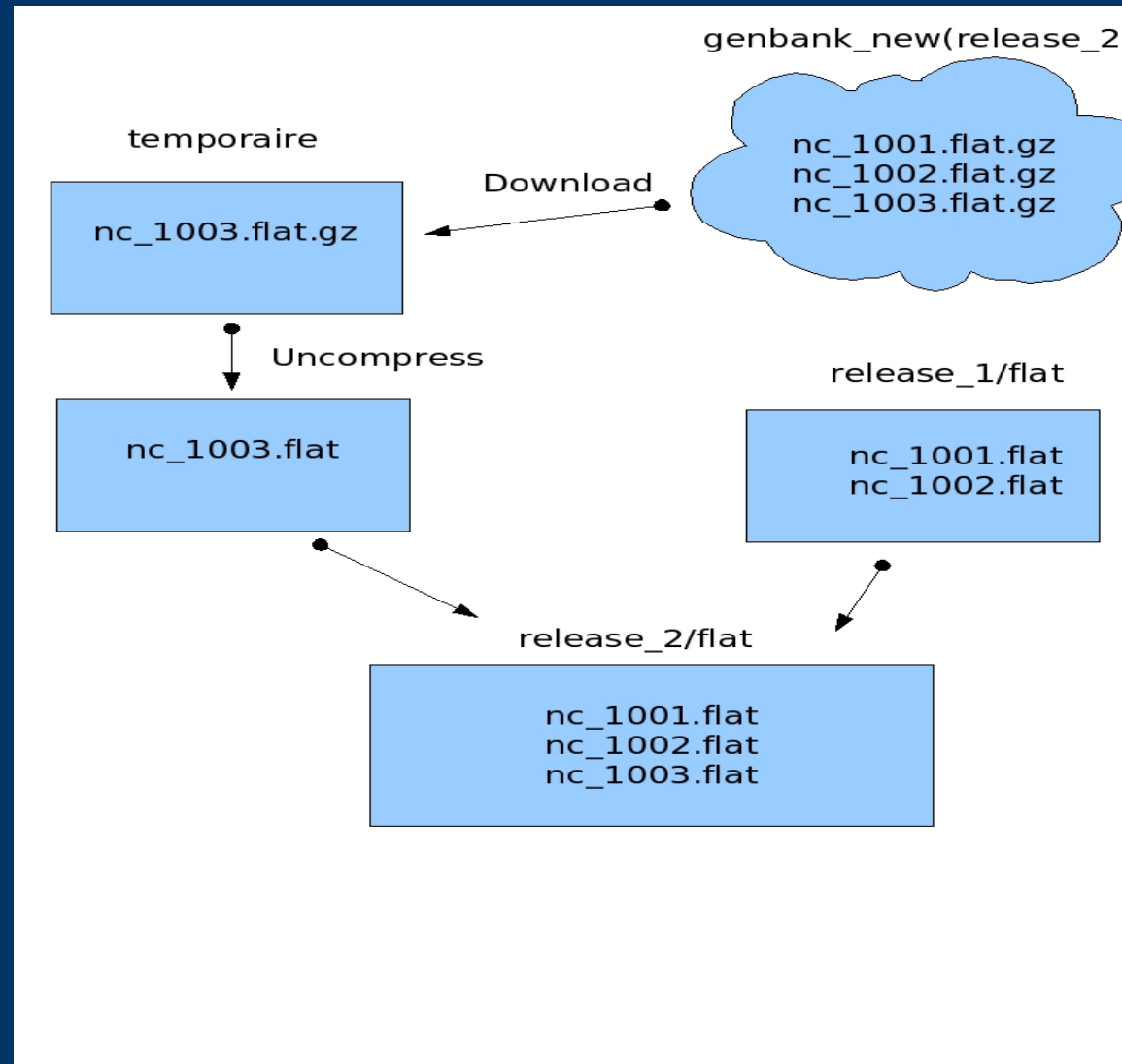
- Application Web (inra toulouse)
-
-

Points forts de Citrina

- **Moteur de workflow de maintenance de banque biologique**
- **Utilise une techno fiable (ANT)**
- **Outils Open source**
 - Libre redistribution, Travaux dérivés possibles
- **Synchronisation des données**



Synchronisation des données



Points faibles

- Logiciel jeune
 - prototype
- Manque de fonctionnalités



Apport du projet Biomaj

- **Refonte de l'architecture**
 - **Reprise sur erreur**
 - › Application plus robuste, mise en production fiable
 - **Historique/Contexte des banques**
 - › Notion de session, réutilisation des attributs dynamiques d'anciennes sessions
 - **Journalisation des sessions**
 - **Enrichissement fonctionnel du workflow**
 - › Banques virtuelles, dépendance de banques, release
 - **Nouveaux protocoles de communication**
 - › Performances et sécurités (téléchargement)
 - › [http,local,rsync](http://local.rsync)

Acteurs du projet

- Mode collaboratif inter-plateforme
 - Enrichissement fonctionnel du workflow
 - Inria Rennes
 - Comportement des workflows
 - Inra toulouse, Inra jouy, Inria rennes
 - Interface
 - Inra toulouse



Biomaj : Calendrier 2006/2007

- **Décembre 2006**

- Mise en production d'une première version de biomaj sur les 3 sites partenaires (sans traitement des calculs d'indexation)

- **Mars 2007**

- Mise en production de la v1

- **Mai 2007**

- **Début des spécifications de la v2**
Sites partenaires: INRA Toulouse, INRA Jouy en jossas, INRIA Rennes
-
-

Biomaj V2

- **BiomajI**

- Interface Web : Visualisation des données biologiques mis à disposition sur site biomaj

- **Mode fonctionnement en « grille »**

- Echanges des données biologiques entre applications biomaj

- › Evite l'effet « goulet d'étranglement » des téléchargements vers un site de référence.

- › Répartition des calculs des indexations des données



