



# ➔ BioSide : du côté des biologistes

TELECOM Bretagne

&

Station Biologique de Roscoff



## ➔ BioSide : projet à large spectre

### • Couvre l'ensemble d'un processus complexe incluant :

- la description des ressources logicielles (BioDesc)
  - Dédié au bioinformaticien
  - Vise une description de haute qualité du programme
- La fédération des ressources (BioSide<sub>Core</sub>)
  - mise en place une architecture robuste et extensible d'accès à une fédération de ressources
- Les workflows
  - BioSide<sub>IHM</sub> : solution conviviale à la spécification / conservation des workflows
  - BioSide<sub>Core</sub> : centralisation et planification des exécutions de workflows
- L'échange de workflow et le partage des connaissances (BioComm)
  - permet de publier / rechercher les workflows





## ➔ BioSide : projet alternatif

- **Un petit projet**
  - 6-7 personnes à temps partiel sur deux sites
- **Focus qualité**
  - Fiabilité de la plateforme
  - Traçabilité des résultats de workflows
  - Services avancés : forte orientation utilisateurs
- **Des choix de conception précis**
  - Plateforme de fédération adaptée à des architectures de déploiement variés (monoposte, C/S, couplage)
  - Qualité des descriptions de ressources



## ➔ Objectifs de la présentation

- **Présenter les choix de conception de BioSide**
  - Plateforme de fédération
  - Architecture contrôlée et ouverte
  - Description de haut niveau des ressources
- **Faire une démonstration des outils**
- **Présenter le travail en cours et les perspectives**
  - Plateforme
  - Recherche

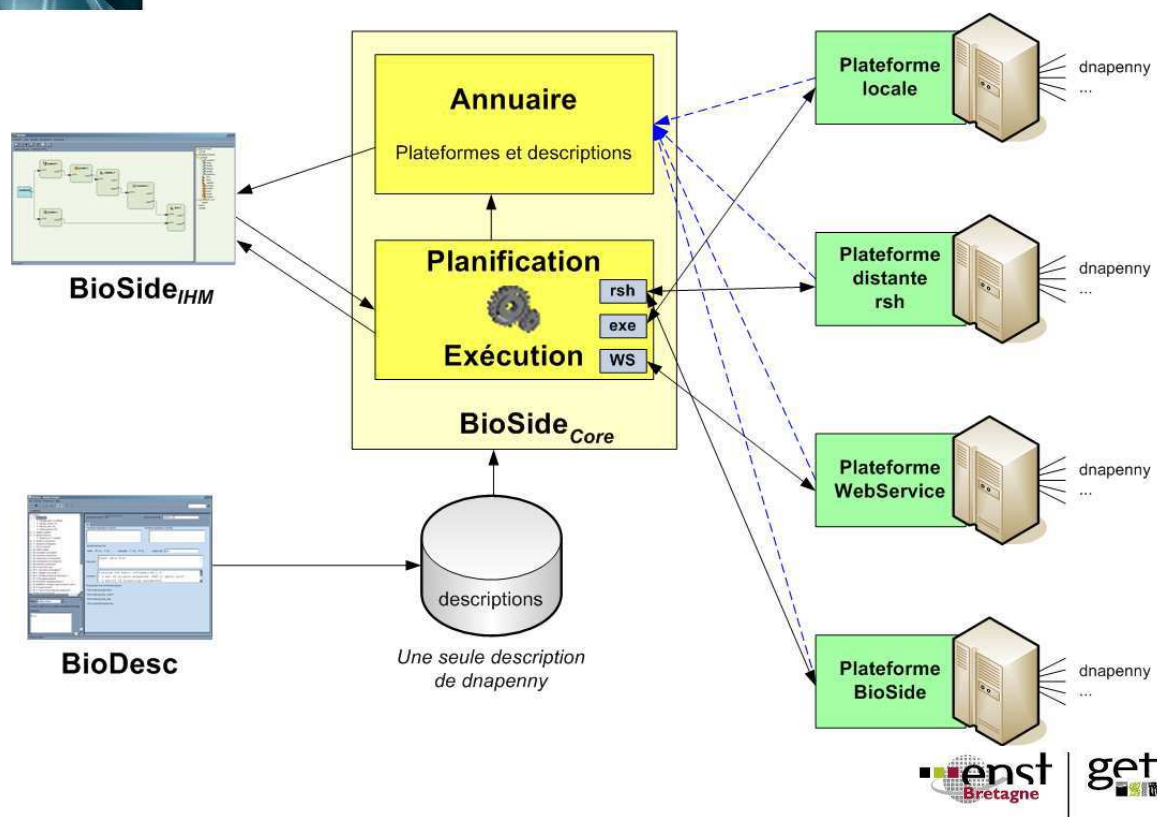


## ➔ Fédération : éléments de choix

- **Fédération de ressources**
  - Annuaire / accès
  - Exigences d'interopérabilité
  
- **En bioinformatique, les ressources sont compliquées**
  - Math et algo
  - Info : nb de paramètres, dépendances, contraintes, formats
  - Ressources informatiques nécessaires
  
- **Combinaison de deux approches**
  - S'appuyer sur des descriptions « complètes »
    - Potentiellement lourd (bioinformaticien fondateur)
  - Ajouter une couche programmée simplifiée et/ou unifiée
    - Risque de multiplication des programmes (BioMoby)



## ➔ Architecture : ouverte et contrôlée





## ➔ Architecture : les choix

- **Diversité des plateformes actuelles**
  - Locales / distantes / protocoles propriétaires ou non en fonction des priorités (facilité d'accès, gros volumes, etc.)
  - Possibilité d'intégrer les autres systèmes via des wrappers ad-hoc
- **Souplesse de déploiement**
  - Possibilité de déployer une solution complète monoposte
  - Possibilité de coupler les plateformes pour obtenir un haut niveau de fédération
- **Moteur d'exécution**
  - Gestion de la redondance, load balancing



## ➔ Description des programmes

- **Les descriptions exhaustives des programmes sont peu valorisées**
  - Elles sont difficiles à réaliser
    - Voir le travail de Pise / ACD
  - Elles doivent être contrôlées
    - Voir les ontologies BioMoby ou les types ACD
  - Les avantages d'une description poussée ne sont pas perçues
    - Permet de contrôler la multiplication des ressources
    - Permet d'offrir des services de haut niveau
- **Comment valoriser les descriptions ?**
  - En proposant un outil dédié à l'expert du programme
  - Support de l'évolution, quel que soit le modèle



## ➔ Démonstration

- **Distribution monoposte**
  - Sélection d'un scénario sur BioCommunity
  - Chargement, examen puis exécution sur BioSide, examen des traces, visualiseurs
  - Modification et exécution sur BioSide
  - Production d'un nouveau scénario et indexation sur Biocommunity
- **BioSide plateforme d'intégration**
  - Utilisation de services Biomoby dans un workflow BioSide



## ➔ Travaux en cours – évaluation

- **Evaluation de l'outil sur le site de Roscoff**
  - Présentation – mise à disposition – assistance – entretien
  - 35 personnes – vaste panel (DR, technicien, bioanalyste)
  - Entre 20 et 30 exécutions par semaine
- **Premières réactions**
  - Réactions globalement positives
    - sur les perspectives liées à l'échange de workflows
    - sur la durée de vie de l'outil
    - Sur le « propriétaire » d'un workflow publié
    - Sur l'indexation d'un workflow
    - Sur le stockage d'un workflow
    - La convivialité semble un facteur discriminant
    - Le manque d'une exécution pas-à-pas et d'une possibilité d'interaction utilisateur
    - Biocommunity.org : utilisent mais ne produisent pas



## ➔ Travaux en cours - évaluation

- **Tests de montée en charge**

- **Protéines**

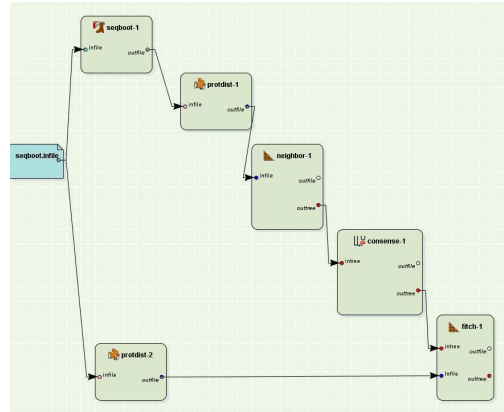
- Humain : 94 séquences
- Mupitetaom : 41 séquences
- Nematost : 91 séquences
- Noalfapdb : 105 séq
- Sigma : 64 séquences

- **Exécution avec N=1000**

- Sortie protdist de 20 à 110 Mo

- **Station Unix**

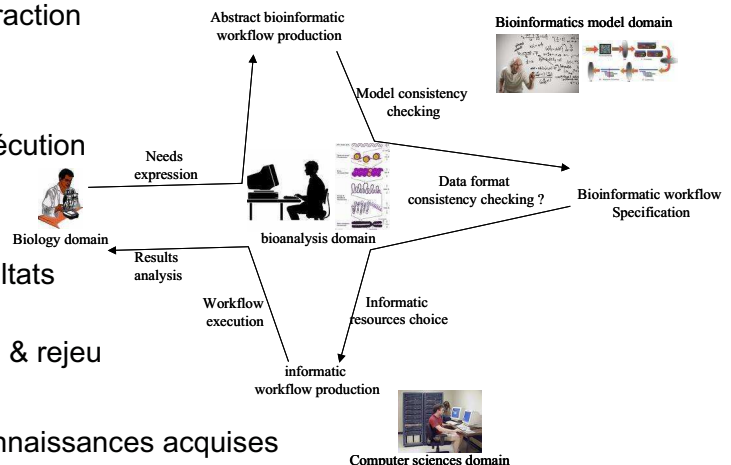
- 10 à 12h de traitement pour les gros jeux de données



## ➔ Travaux en cours : CapExBio

- **Processus coopératif et itératif (spirale)**

- La compréhension du besoin
- La recherche / proposition d'une approche
- Recherche de ressources adaptées
  - Nombre, pertinence, abstraction
  - Qualité des descriptions
- Contrôles
  - De la conception et à l'exécution
  - Qualité de la plateforme
- Traçabilité
  - De l'exécution et des résultats
  - Plateformes multiples
  - Possibilité de modification & rejou
- Publication
  - Mise à disposition des connaissances acquises



## ➔ Conclusion et perspectives

- **Premier bilan**
  - L'architecture est stabilisée
  - Première évaluation en cours (positive pour le moment)
- **Prévisions**
  - Distribution C/S avec Web Services
- **Une coopération contrôlée**
  - Production de descriptions
  - Production de workflow
  - Production de connaissances
  - Cycle vertueux ?



## ➔ Contacts

- **TELECOM Bretagne**
  - Philippe Picouet [Philippe.Picouet@telecom-bretagne.eu](mailto:Philippe.Picouet@telecom-bretagne.eu)
  - Sébastien Bigaret
  - Philippe Tanguy
  - Frédéric Cadier
- **Station Biologique de Roscoff**
  - Erwan Corre [corre@sb-roscoff.fr](mailto:corre@sb-roscoff.fr)
  - Xavier Bailly [bailly@sb-roscoff.fr](mailto:bailly@sb-roscoff.fr)
  - Gildas Le Corguillé
  - Et Olivier Collin, ex SBR et sympathisant.

