



Portails Chêne, Peuplier et Pin – GnpIS - Evoltree

un exemple de systèmes d'information interoperables

François Ehrenmann, Célia Michotey, Véronique Jorge, Catherine Bastien, Odile Rogier, Johanna Schmidt, Stephan Gaubitzer

AG CATI SIOEA, 20 Mai 2015, Montpellier, France



Portails Chêne, Peuplier et Pin – GnplS - Evoltree communiquer sans diviser

Portails Chêne, Peuplier et Pin

- › Ressources génétiques et génomiques (*séquences, marqueurs, diversité, ...*)
- › 10 bases de données / applications web

GnplS

- › Système d'information **institutionnel** et de **référence** multi-espèces dédié aux plantes et bio-agresseurs (phénotypes, polymorphismes, marqueurs, séquences, ...)
- › Système modulaire : outils spécifiques développés localement + outils open source + bases de données / application web

Evoltree eLab

- › Moteur de recherche centralisé
- › Base de données / application web

→ **Quelle communication ? Comment ?**

Portails Chêne, Peuplier et Pin

Systemes d'information pour les données génétiques et génomiques

Portail Chêne

- › Informations générales sur la famille des *Fagaceae*
- › Ressources génétiques et génomiques

Portail Pin (Décembre 2012)

- › Informations générales sur la famille des *Pinaceae*
- › Ressources génétiques et génomiques

Portail Peuplier (en cours de développement)

- › Informations générales sur la famille des *Salicaceae*
- › Ressources génétiques et génomiques

→ **10 bases de données / applications web**

Portails Chêne, Peuplier et Pin

Hébergement et contenu des bases de données

INRA Pierroton

Quercus Map
Pinus Map
Populus Map
single tree genotypic and phenotypic data

Oak provenance
passport data of provenances

SSR
microsatellites

CMap
maps and QTL

GD²
georeferenced database of genetic diversity

INRA Orléans

Treepop
Genotypic, phenotypic data of association populations

INRA Toulouse

ESTs
EST sequences

Institut technologique d'Autriche

SNP
Single Nucleotide Polymorphisms in candidate genes of oaks

Candidate genes

- Hétérogénéité des systèmes, structurelle et sémantique
- Hétérogénéité des SGBD et des applications web
- Hétérogénéité des ontologies (taxonomie, traits, coordonnées, populations...)

Interopérabilité Evoltree - Portails

Développement d'un moteur de recherche intégré

- Un seul point d'accès doit pouvoir accéder à l'ensemble des bases de données
- Idée principale : s'affranchir des différences entre les bases de données en utilisant des attributs standardisés pour chaque « item » ou élément

Exemple : les arbres sont stockés de différentes manières et avec différents attributs dans plusieurs bases de données.

- Comment accéder aux données ?



Interopérabilité Evoltree - Portails

Développement d'un moteur de recherche intégré

- Base de données centrale

Interopérabilité Evoltree - Portails

Développement d'un moteur de recherche intégré

- ~~Base de données centrale~~
- Accès direct aux bases de données

Interopérabilité Evoltree - Portails

Développement d'un moteur de recherche intégré

- ~~Base de données centrale~~
- ~~Accès direct aux bases de données~~
- Web services

Interopérabilité Evoltree - Portails

Développement d'un moteur de recherche intégré

- ~~Base de données centrale~~
- ~~Accès direct aux bases de données~~
- ~~Web services~~

➤ **PHP interface (TAPIR)**

The screenshot displays the TAPIR interface for configuring a data source. The 'Mapping' tab is active, showing a list of available schemas to map. The selected schema is 'PICME - Evoltree Libs'. Below the schema list, there are buttons for 'unmap', 'automap', and 'is unmaped'. A table defines the mapping between concepts and database fields:

concept	searchable	column	table	type	
(*) id	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	id	numeric
(?) cnsl	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	cnsl_id	text
(?) species	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	species	text
type	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	type	text
organ	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	organ	text
tissue	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	tissue	text
dev_stage	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	dev_stage	text
comments	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	comments	text
owner	<input checked="" type="checkbox"/>	single column	v_evoltree_libs	owner	text

At the bottom, there is a 'save changes' button.

Description du protocole TAPIR

Protocole

- › Protocole de requête/réponse pour accéder à des données structurées
- › Données et bases de données indépendant

Implémentation

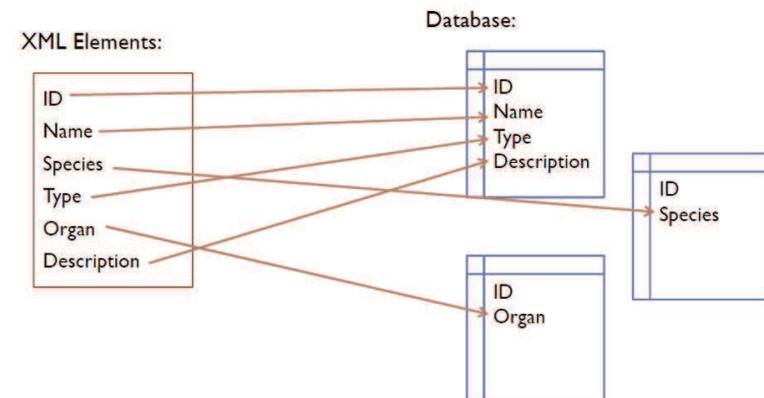
- › PHP 
- › Open source

Interface

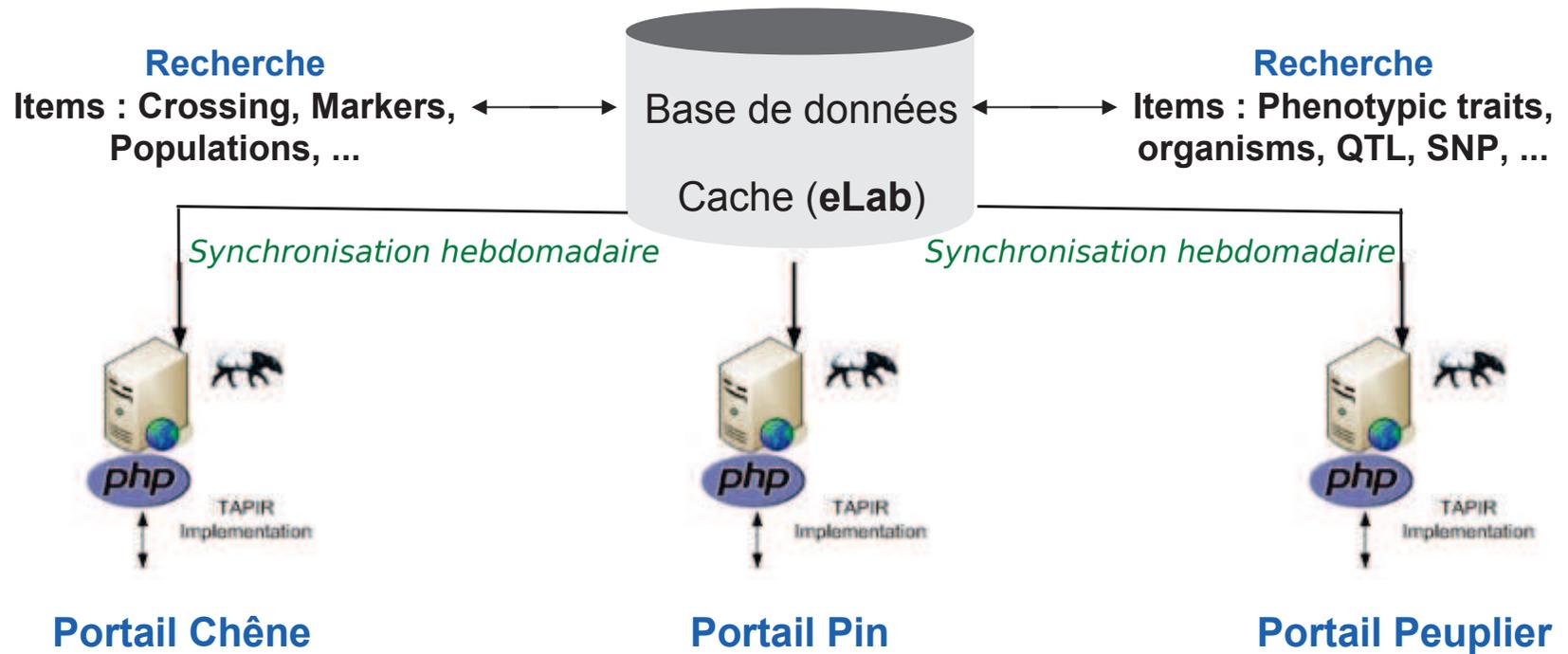
- › Serveur web + PHP accessibles
- › Pas de librairies additionnelles

Transfert

- › Donnée transférées via HTTP
- › Format XML définit la structure des données **XML**
- › Mapping définit entre éléments XML et champs de la base

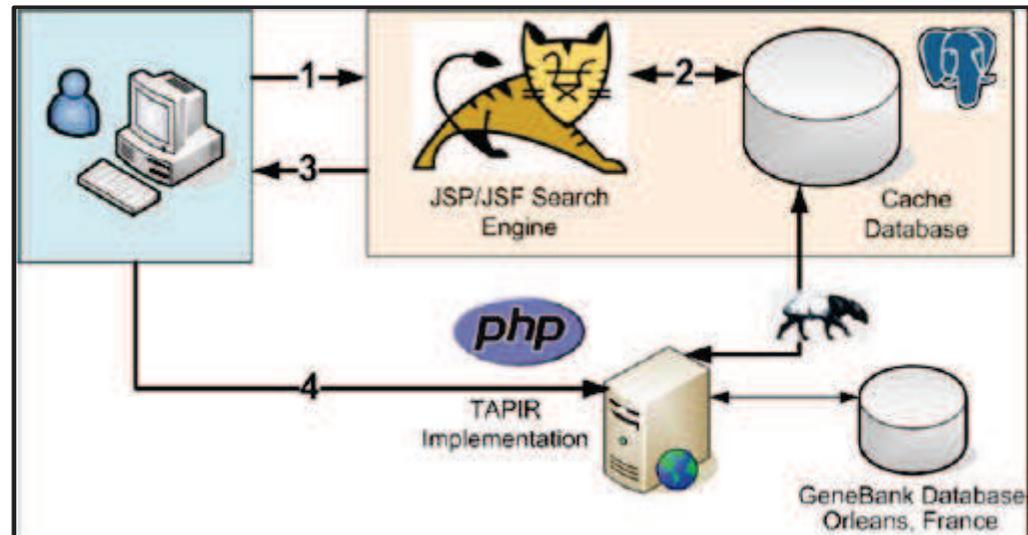


Evoltree - Portails : schéma



EvoTREE – Portails

Exemple de requête



1. L'utilisateur crée sa requête sur l'interface elab (Ex : Trouver toutes les informations sur toutes les espèces d'arbres ayant été utilisées pour réaliser des cartes génétiques)
2. Le moteur de recherche effectue la recherche dans la base de données « cache »
3. Les résultats de la recherche sont renvoyés à l'utilisateur
4. Si l'utilisateur veut obtenir plus de détails sur un résultat, il a la possibilité de cliquer sur un lien de redirection vers la base externe qui contient le résultat à son origine. Ses informations de connexion (login, password) sont envoyés à la base externe et le résultat est affiché via l'interface connectée à cette base externe

Interopérabilité Evoltree - Portails

Le Elab Evoltree permet :

- Aux utilisateurs d'interroger la base de données « centrale »
- De stocker uniquement les données nécessaires pour la définition de jeux de données
- De rediriger les utilisateurs vers les bases de données partenaires afin d'obtenir de plus amples informations

Details for Association Population:

Name:	Drome6	Available in:	TreePop Database
Species:	Populus nigra	Owner:	Marc Villar, INRA Website
Description:			
Size:	63 Individuals		

Location: Individuals Phenotypic Traits Phenotypic Traits (Individuals)

Taxonomy

Genus:

- Lymantria (9)
- Malus (2)
- Picea (2276)
- Pinus (39778)
- Pisolithus (1)
- Pisum (13)
- Podocarpus (1)
- Populus (20513)

Species:

- Populus deltoides x Populus maximowiczii (1)
- Populus deltoides x Populus nigra (231)
- Populus deltoides x Populus trichocarpa (3044)
- Populus maximowiczii (3)
- Populus maximowiczii x [Populus deltoides x P
- Populus maximowiczii x Populus nigra (1)
- Populus maximowiczii x Populus trichocarpa (1)
- Populus nigra (7458)

Data Owner

Organisation:

INRA (5098)

DataItems

Category: Population (22)

Datatype: -- Not selected --

Additional Parameters

Genetic/Phenotypic Data: -- Not selected --

Country: France (7), Germany (1)

Adm. Level 1: Alsace (3), Bourgogne (1), Centre (2), Rhône-Alpes (1)

Adm. Level 2: -- Select an area 1 first --

7 DataItems found:

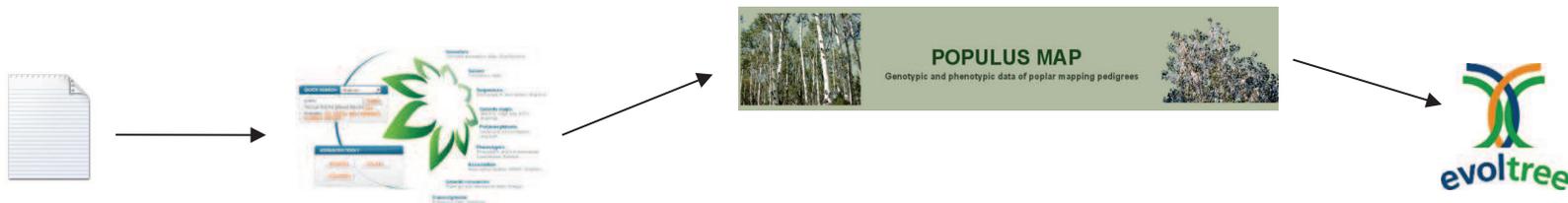
Type	Name	Info	Species
1	Drome6	Association population with 63 individuals	Populus nigra
2	BSL	Association population with 42 individuals	Populus nigra

Interopérabilité GnpIS – Portails – Evoltree

Données Peuplier pour la mise en place d'un flux

Flux d'alimentation données locales Orléans -> GnpIS -> Portail salicaceae -> Evoltree

- Alimentation **GnpIS** par des fichiers plats formatés au format de soumission Siregal (validation manuelle)
- Alimentation du **portail peuplier** à partir de GnpIS via scripts automatiques d'extraction de données publiques de peuplier
 - Création de fichiers temporaires au format prédéfini sur un espace d'échange sur les serveurs de l'URGI
 - Récupération fichiers (scp) et insertion/MAJ des données (étape automatisée, fréquence déterminée)
- Alimentation du **eLab evoltree** via le protocole TAPIR vu précédemment (étape automatisée, fréquence déterminée)

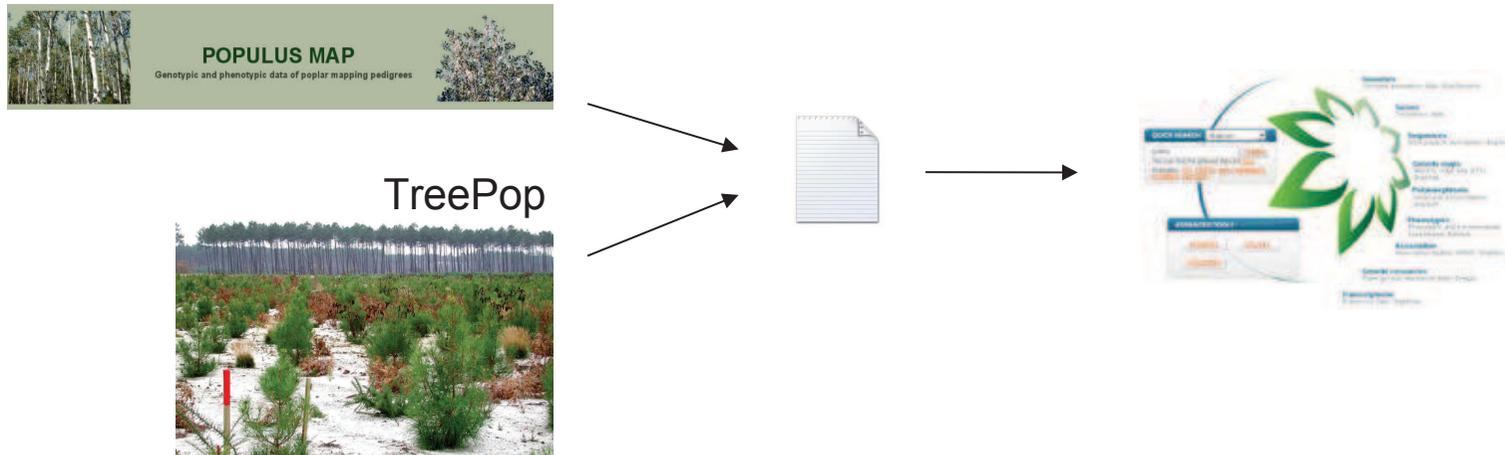


Interopérabilité GnpIS – Portails – Evoltree

Données Peuplier pour la mise en place d'un flux

Flux d'alimentation Portail salicaceae -> GnpIS

- Extraction de données du portail au format Siregal
- Comparaison données soumises dans GnpIS et bases du portail
- Homogénéisation des accessions, de la taxonomie, de la généalogie et des collections



Interopérabilité GnpIS – Portails – Evoltree

Données Peuplier pour la mise en place d'un flux

Formats de soumission, Thésaurus et interactions

- › Divers formats de soumission « enrichis » pour pouvoir gérer au mieux les différents types de données (populations, collections, familles, etc...)
- › Travail d'homogénéisation de l'ontologie utilisée par GnpIS afin de tenir compte des termes utilisés dans les différents SI - Rapprochement souhaité avec des ontologies internationales pour la gestion multi-langues
- › Recensement des URL à utiliser pour une bonne communication entre applications web / bases de données (numéros d'accession et/ou taxon)

Populus deltoides x Populus lasiocarpa	Peuplier Poplar	Eukaryota	Plantae	Magnoliophyta	Magnoliopsida	Malpighiales	Salicaceae	Populus	W. Bartram (1785)	
Populus deltoides x Populus maximowiczii	Peuplier Poplar	Eukaryota	Plantae	Magnoliophyta	Magnoliopsida	Malpighiales	Salicaceae	Populus	W. Bartram (1785)	
Populus deltoides x Populus nigra	Peuplier Poplar	Eukaryota	Plantae	Magnoliophyta	Magnoliopsida	Malpighiales	Salicaceae	Populus	W. Bartram (1785)	
Populus deltoides x Populus trichocarpa	Peuplier Poplar	Eukaryota	Plantae	Magnoliophyta	Magnoliopsida	Malpighiales	Salicaceae	Populus	W. Bartram (1785)	
Populus euphratica	Desert poplar Euphrates poplar	Eukaryota	Plantae	Magnoliophyta	Magnoliopsida	Malpighiales	Salicaceae	Populus	Oliv.	

Base de données PopulusMap

Taxons *GnpIS taxons*

The query has been restricted on: [Help!](#)

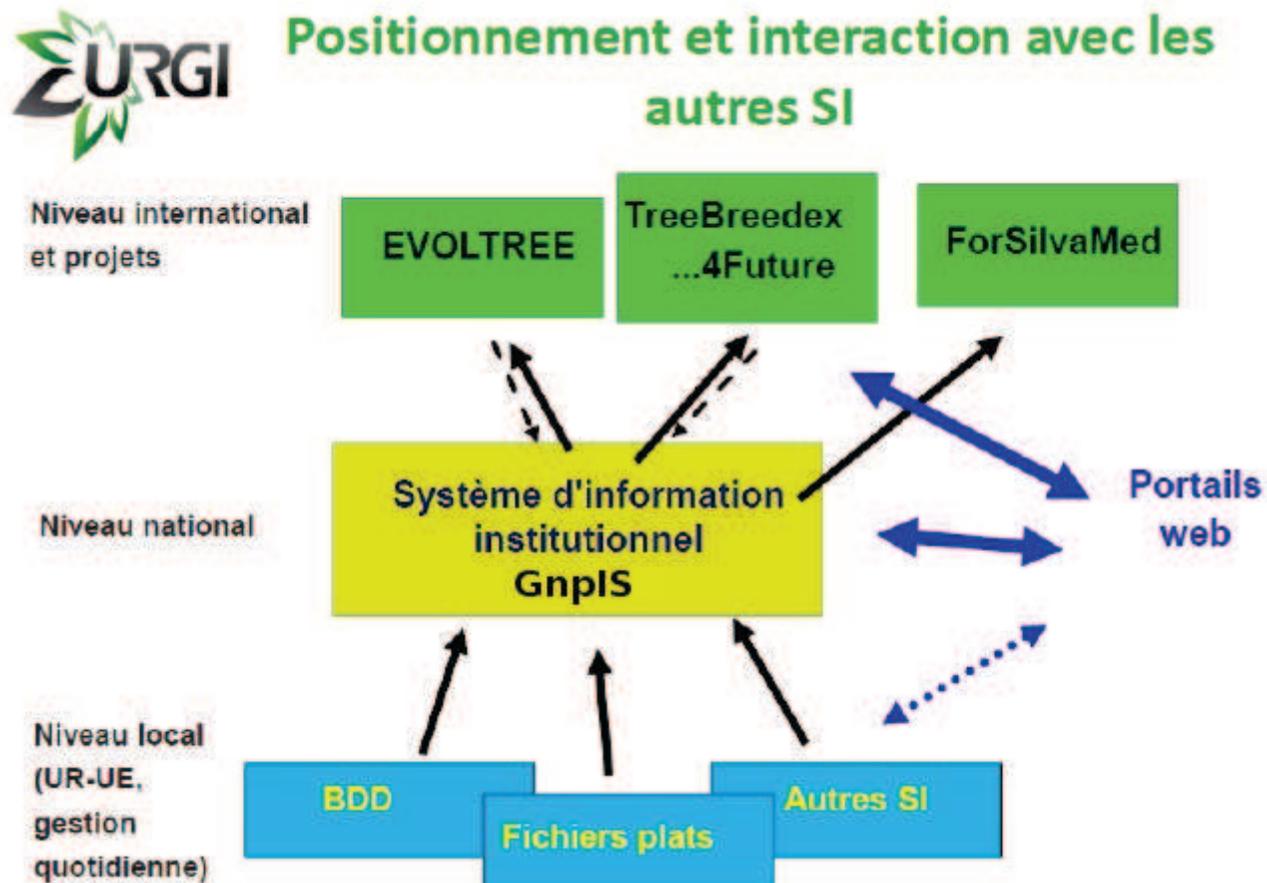
Query parameters:
 Scientific name: Populus deltoides
 Thematic restriction: accessions

6 items found, displaying 1 to 8 | Display 10 results per page

#	Scientific name	Common names	Genetic maps	Polymorphism experiments	NGS experiments	Expression	Genplastids	Phenotypic traits
1	<i>Populus deltoides</i>	Match poplar, Eastern cottonwood, Liard, Peuplier de Virginie, Peuplier deltoïde, Peuplier noir d'Amérique	-	-	1	-	25	-
2	<i>Populus deltoides</i> x <i>Populus lasiocarpa</i>	Peuplier, Poplar	-	-	-	-	-	-

Positionnement et interactions entre SI

Pour le département EFPA, le positionnement et l'interaction entre les différents Système d'Information (SI) se schématise ainsi :



Services interopérables opérationnels

- Portails Chêne et Pin - Evoltree

- Le **eLab** permet une recherche intégrée sur toutes les bases de données des portails actuellement disponibles dans le projet de EVOLTREE.
- Multi-plateforme, différents langages et SGBD
- Protocole **TAPIR** (XML & PHP)
- Format d'échange XML

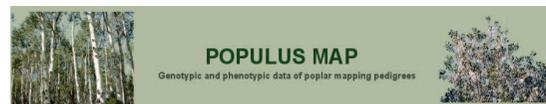
- GnpIS - Portail Peuplier

- Flux de données publiques formatées (accès sécurisé)
- Plateforme UNIX, SQL
- Echange de fichiers à plat

Services interoperables à venir

- Portail Peuplier – Evoltree

- Permettre au **eLab** la recherche intégrée sur les données **Peuplier**
- Intégrer dans le portail peuplier les ressources génétiques, les données de cartographie, phénotypage, génotypage
- Protocole **TAPIR** (XML & PHP) à implémenter





Merci de votre attention

