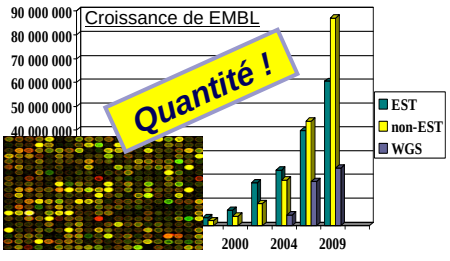
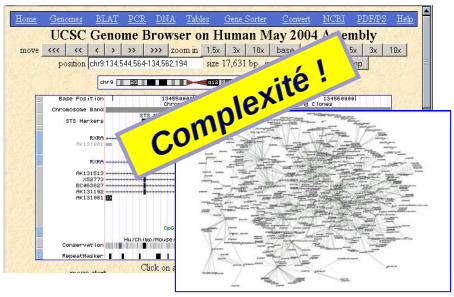
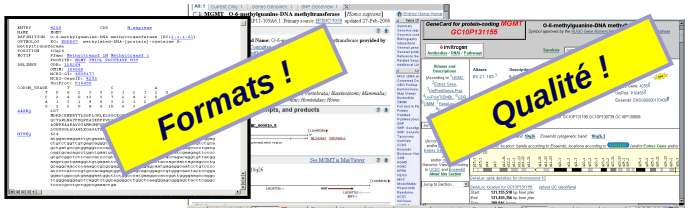
Entrepôts de données personnalisés pour l'identification de gènes candidats - Construction de jeux de données intégrés, guidée par les connaissances du domaine

Marie-Dominique Devignes
Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA)
Equipe Orpailleur
Nancy

Intégration de données hétérogènes en biologie







Paradoxe : Trop d'info tue l'info

Multiplicité des bases de données biologiques

- Bref historique

  1965 : Atlas of Protein Sequences and Structures : Margaret Dayhoff
  - ***→ PIR en 1986 (Université Georgetown)***

  1980's : EMBL/GenBank/DDBJ ; SwissProt

  2K's : Compilations Nucleic Acids Research (Janvier)
  - ***Janv 2003 : > 300 BD***
  - ***Janv 2010 : > 1200 BD***

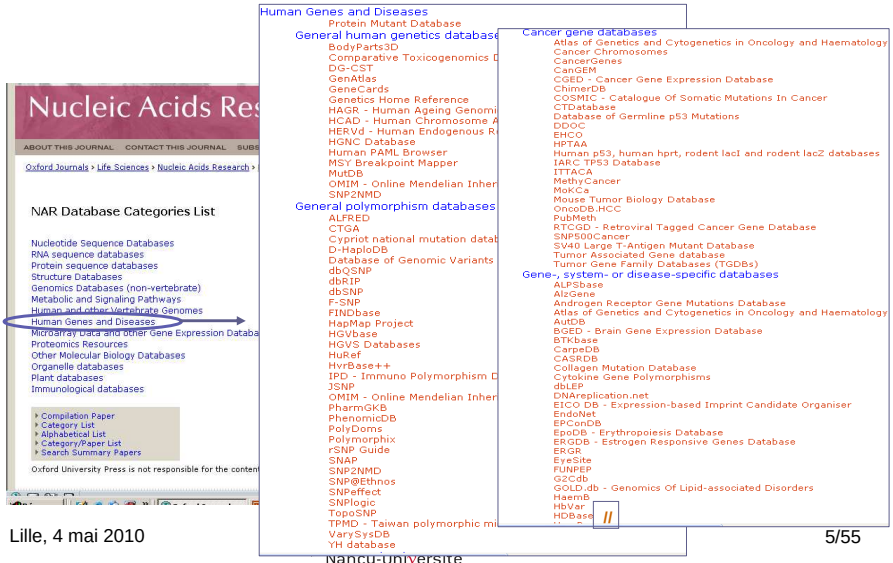
- Problématique de l'intégration de ces bases de données
  - **Dès 1994-1995 le problème est soulevé**
    - ***P Karp, D Markowitz, S Davidson***
    - ***L. Stein, Nature 2002 - > 2006 « A bioinformatic nation » « Cyberinfrastructure »***
    - ***R Altman 2005 Concept de « Ressourceome »***
  - **Workshops: DILS, International Symposium on Integrative Bioinformatics, etc.**

Motivation: la recherche « in silico » ….

- Double problème :
  - **Pb1 : Choix des bases de données (BD) à interroger/explorer**
  - **Pb2 : Intégration des données collectées**
- Exemple :

« Parmi les 50 gènes de ce locus remanié sur le chromosome X, lesquels sont exprimés dans le cerveau et annotés par des termes GO en rapport avec la migration neuronale »
  - **Pb1 : Recherche de BDs**
    - ***Sur l'expression des gènes***
    - ***Sur l'annotation des gènes***
    - ***Sur les gènes du chromosome X ou leurs orthologues dans d'autres espèces***
  - **Pb2 : Intégration des données collectées**
    - ***Mise en relation des données provenant de BD distinctes***
    - ***Détection possible d'incohérences***

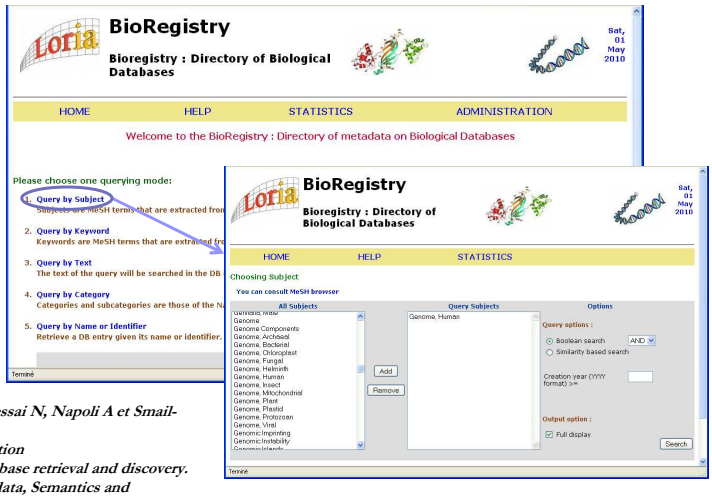
Pb1 : Choix des bases de données



Nancy-Université

Exemple d'annuaire interrogeable en ligne

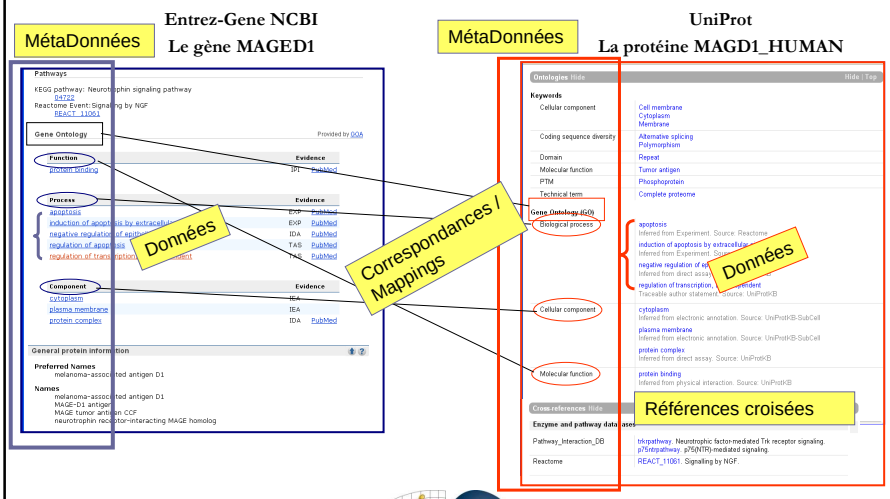
Les BD du catalogue NAR sont indexées automatiquement par des métadonnées « Subject » issues du vocabulaire MeSH



Devignes MD, Franiatte P, Messai N, Napoli A et Smail-Tabbone M (2010)
BioRegistry : automatic extraction of metadata for biological database retrieval and discovery.
International Journal on Metadata, Semantics and Ontologies (sous presse).

Nancy-Université

Pb2 : Intégration de données hétérogènes



Pb2 : Intégration de donnnées

- Différents niveaux
  - **Données**
  - **Métadonnées**
  - **Correspondances / Mapping**
  - **Références croisées**
- Rôle des connaissances
  - **Comparaison des données**
    - ***Exemple des annotations fonctionnelles : rôle des vocabulaires, mesures de similarité sémantiques***
  - **Signification des Métadonnées**
    - ***Schéma, Indexation***
  - **Correspondances / Mappings**
    - ***Connaissances de l'expert***

PLAN DE L'EXPOSÉ

I. Les différentes stratégies pour l'intégration de données

II. ACGR : Un entrepôt de données personnalisé pour l'identification de gènes candidats

III. MODIM : Généralisation à la construction de jeux de données pour la fouille de données

PLAN DE L'EXPOSÉ

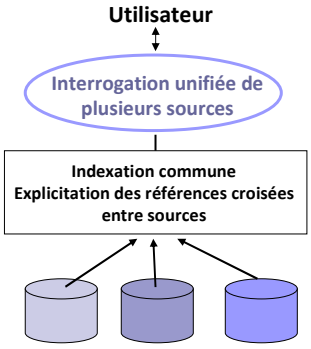
I. Les différentes stratégies pour l'intégration de données
   1. **Essai de classification**
   2. **Avantages-Inconvénients de quelques systèmes fonctionnels**
      - ***Système d'interface unifiée***
      - ***Base de données intégrée***
      - ***Workflows et web services***
   3. **Un système « hybride : l'entrepôt personnalisé**

II. ACGR : Un entrepôt de données personnalisé pour l'identification de gènes candidats

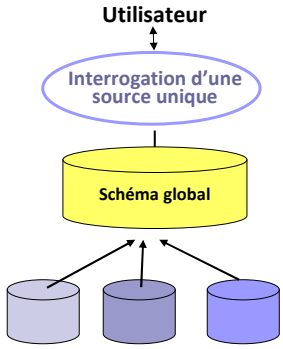
III. MODIM : Généralisation à la construction de jeux de données pour la fouille de données

1. Essai de classification (1)

Avec ou sans schéma intégrateur



1 :Systèmes multi-bases unifiant l'accès (croisé) à plusieurs sources (ACNUC, Entrez, SRS, BioMart…)



2 : Deux types de systèmes intégrés
Matérialisé = BD unique, entrepôt (GeneCards, InterPro, BioWareHouse…)
Non-matérialisé = Médiateur (TAMBIS …)

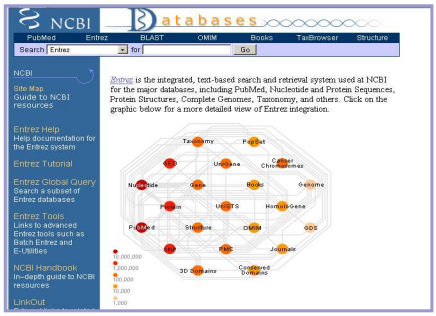
1. Essai de classification (2)

Avec flots de traitements et/ou visualisation

- Enchaînement d'outils et d'accès aux bases de données
  - Accès standardisés par la technologie des services web
  - Notion de workflow
  - Exemple Taverna (UK MyGRID)
- Plateforme d'intégration et de visualisation
  - Exemple ONDEX (représentation à base de graphes)

2. Avantages-Inconvénients
2.1 Interface d'interrogation unifiée

- Exemple Entrez-NCBI
  - **Entrez: ensemble des BD de biologie moléculaire du NCBI + un système d'interrogation particulier associé à ces banques**

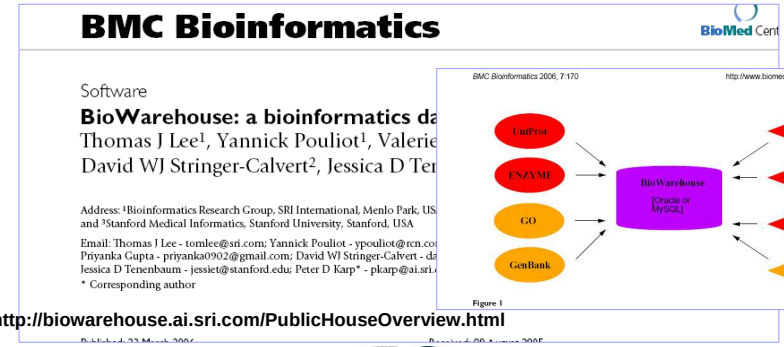


2. Avantages-Inconvénients
2.1.Interface d'interrogation unifiée (suite)

- Avantages
  - Préserve l'autonomie des BD
  - Utilisation facile, intuitive
  - Références croisées entre les BD
  - Exportation dans divers format
    - *Permet la constitution d'un jeu de données personnelles*
  - Existence de systèmes ouverts, personnalisables en terme de nombre et nature des BD interfacées (SRS, BioMart)
- Limites
  - NCBI-Entrez : système fermé, seulement les BD du NCBI
  - SRS : Maintenance, nécessité de ré-indexer régulièrement pour tenir compte des mises à jour
  - Pas d'exploitation des relations entre les données elles-mêmes
  - Connaissances requises a priori sur les BD et leur qualité

2. Avantages-Inconvénients
2.2 Base de données intégrée / Entrepôts

- Exemple BioWarehouse
  - **Système d'entrepôt configurable permettant d'importer des BD entières et de gérer les données localement selon un schéma global simple et extensible**


BMC Bioinformatics

BioMed Cent

Software

BioWarehouse: a bioinformatics da

Thomas J Lee[1], Yannick Pouliot[1], Valerie

David WJ Stringer-Calvert[2], Jessica D Ter

Address: [1]Bioinformatics Research Group, SRI International, Menlo Park, US and [2]Stanford Medical Informatics, Stanford University, Stanford, USA

Email: Thomas J Lee - tomlee@sri.com; Yannick Pouliot - ypouliot@rcn.co Priyanka Gupta - priyanka0902@gmail.com; David WJ Stringer-Calvert - da Jessica D Tenenbaum - jessiet@stanford.edu; Peter D Karp* - pkarp@ai.sri.

* Corresponding author

Figure 1

http://biowarehouse.ai.sri.com/PublicHouseOverview.html

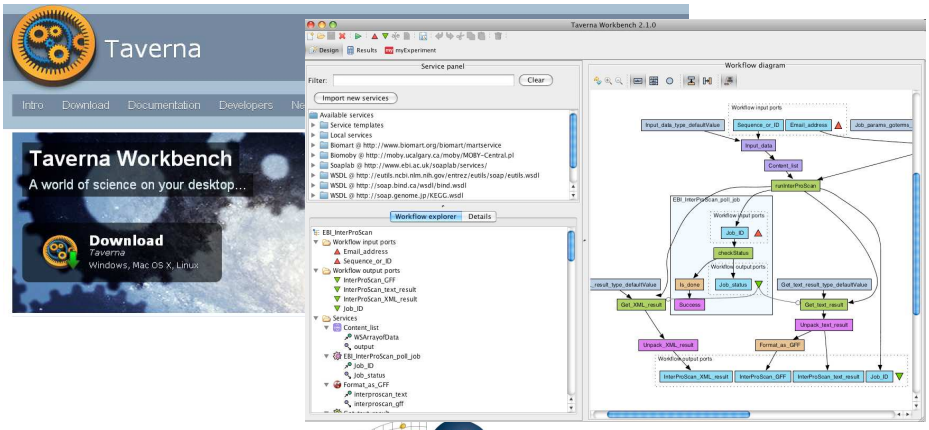
2. Avantages-Inconvénients
2.2 Base de données intégrée / Entrepôts (suite)

- Avantages
  - **Robustesse**
  - **Rapidité d'accès**
  - **Volumes traités**
  - **Nettoyage des données, résolution des conflits entre les données**
  - **L'utilisateur n'a pas besoin de connaître les détails des modèles de données de chacune des BD sources**

- Inconvénients
  - **Problème des mises à jour : repeupler l'entrepôt**
  - **Complexité du système**
  - **Coût de l'ajout d'une nouvelle BD (développements nécessaire, adaptation du schéma global)**

2. Avantages et Inconvénients
2.3 Plateforme de « workflows »

- Exemple Taverna MyGrid **http://www.taverna.org.uk/introduction/what-is-taverna/**



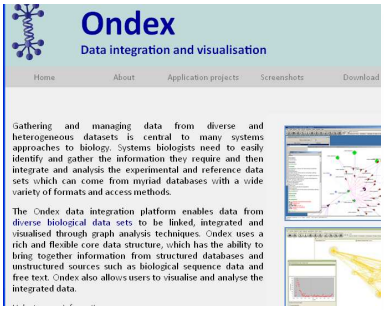
2. Avantages et Inconvénients
2.3 Plateforme de « workflows » (suite)

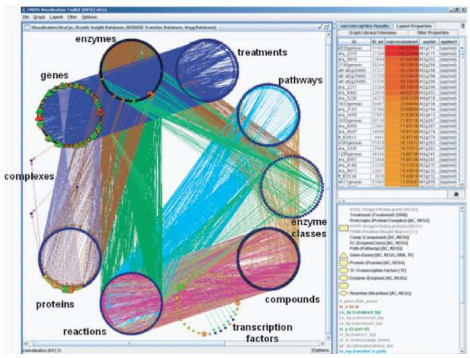
- Avantages
  - **Exécutions répétitives d'enchaînements d'étapes**
  - **Gain de temps**
  - **Possibilité de partage des workflows**
  - **Flexibilité (plus de 1600 services web accessibles)**
- Inconvénients
  - **Prise en main relativement lourde**
  - **Pas d'intégration des résultats des traitements**
    - ***Travail d'analyse postérieur à l'exécution des workflows***
- Repris par d'autres exposés de la journée

2. Avantages-Inconvénients

2.4 Plateformes d'intégration et de visualisation

- Exemple ONDEX
  - **Combiner intégration de BD et fouille de texte grâce à des méthodes d'analyse à base de graphes**





Köhler et al., Bioinformatics 2006/

2. Avantages-Inconvénients

2.4 Plateformes d'intégration et de visualisation (suite)

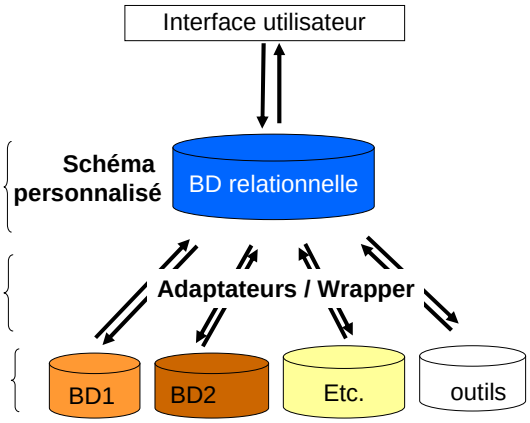
- Avantages
  - **Prise en compte de la visualisation**
  - **Puissance de calcul**
  - **Lien avec les technologies du web sémantique (graphes RDF)**
- Inconvénients
  - **Prise en main relativement lourde**
  - **Coût d'adaptation à un problème d'intérêt difficilement évaluable**

3. L'approche « entrepôt personnalisé »

Une solution hybride et flexible à l'échelle du laboratoire

Avantages et motivation

- Modélisation des données en fonction d'un problème posé
- Robustesse d'une BD relationnelle
- Wrappers génériques
- Notion de scénario
- Choix des BD ouvert
- Respect de l'autonomie des sources (mises à jour)



PLAN DE L'EXPOSÉ

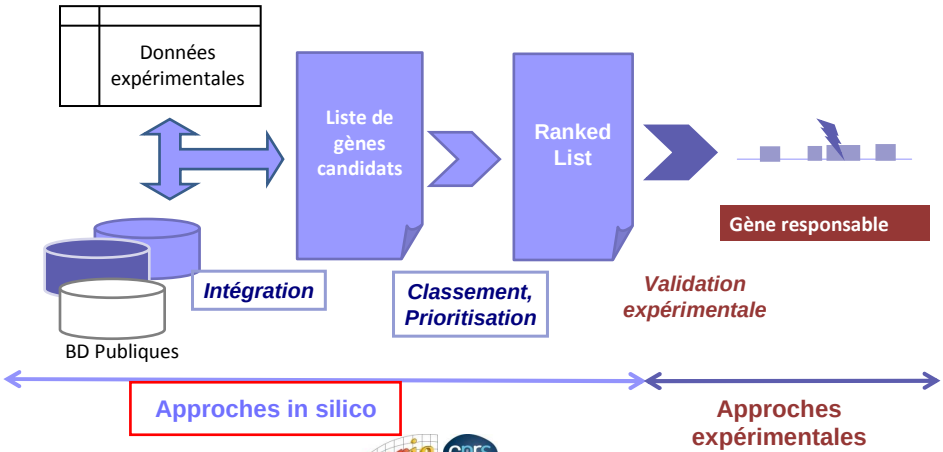
I. Les différentes stratégies pour l'intégration de données

II. ACGR : Un entrepôt de données personnalisé pour l'identification de gènes candidats
1. **La question posée et sa modélisation**
2. **L'approche ACGR - Réalisation**
3. **Résultats pour 2 syndromes rares**

III. MODIM : Généralisation à la construction de jeux de données pour la fouille de données

1. La question posée :

Identifier des gènes candidats pour une maladie donnée



La notion de gène candidat (1)

- Réalité intuitive pour le biologiste
  - Définition du Webster Medical Dictionary
    - ***Any gene thought likely to cause a disease. The gene may be a candidate because it is located in a particular chromosome region suspected of being involved in the disease or its protein product may suggest that it could be the disease gene in question.***
- Difficile à appréhender pour l'informaticien
  - ***Likely to cause a disease***
  - ***Region suspected of being involved,***
  - ***may suggest,***
  - ***it could be the disease gene…***

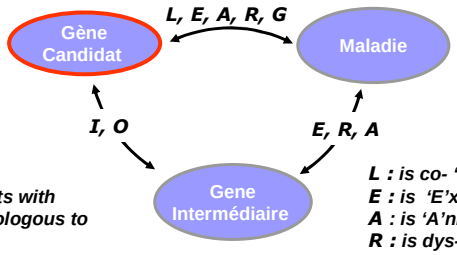
La notion de gène candidat (2)
Essai de formalisation

- Un gène candidat pour une maladie est un gène qui est en relation avec une maladie
  - ***Est localisé dans la région chromosomique liée à la maladie***
  - ***Est exprimé dans le tissu affecté par la maladie***
  - ***Est induit ou réprimé chez les patients atteints de la maladie***
  - ***Est annoté de façon similaire à la maladie***
- Le gène candidat peut aussi être un gène qui est en relation avec un autre gène qui lui est en relation avec la maladie
  - ***Notion de gène intermédiaire***
  - ***Gène orthologue dans un organisme modèle***
  - ***Gène en interaction***

Modélisation ACGR* (1)
Définition des gènes candidats

*ACGR:
Approach for Candidate Gene Retrieval
Yilmaz S et al., Bioinformatics 2009

→ ***Combinatoire de relations gène-maladie et gène-gène***



I *: 'I'nteracts with*
O *: is 'O'rthologous to*

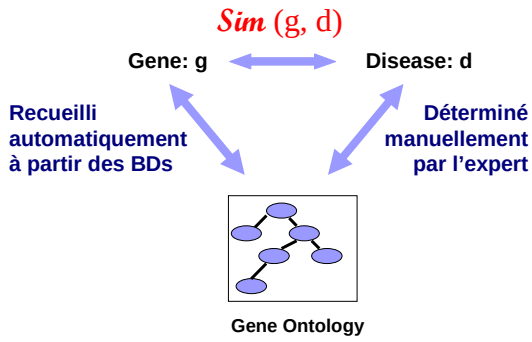
L *: is co- 'L'ocalized on the genome*
E *: is 'E'xpressed in affected tissues*
A *: is 'A'nnotated in a similar way*
R *: is dys- 'R'egulated in affected patients*
G *: is re- arran'G'ed in affected genomes*

→ ***Mise en œuvre des connaissances du domaine, rôle de l'expert du domaine***

Modélisation ACGR (2)

Mesure de similarité

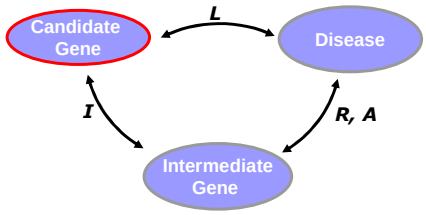
→ *Similarité « sémantique » entre un gène et une maladie à partir des annotations fonctionnelles issues d'un même vocabulaire*



Modélisation ACGR (3)

Résultats attendus

- Tableaux présentant les gènes selon les définitions possibles de gènes candidats, classés selon la similitude avec la maladie
  - **Ex : Liste des gènes co-localisés avec la maladie et interagissant avec un gène intermédiaire dérégulé chez les patients malades / ou annotés de façon similaire à la maladie.**



Modélisation ACGR (4)

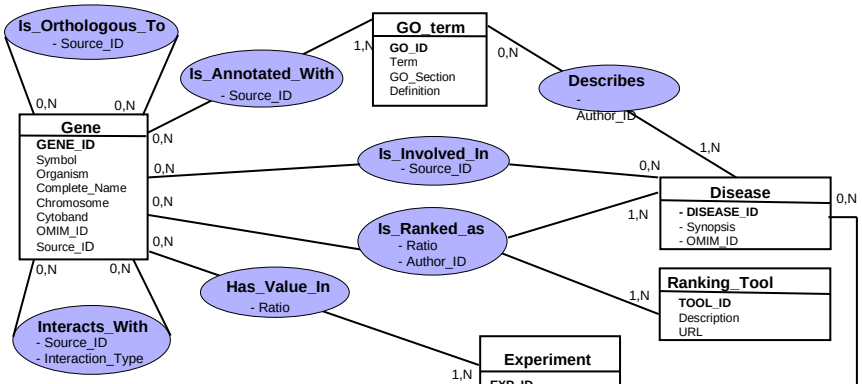
Inventaire des données à collecter et des sources à explorer

Données à collecter	Choix des sources
Annotations GO de la maladie	OMIM + travail de l'expert
Gènes (homme, souris, droso)	NCBI GENE, MGD, Flybase
Localisation chromosomique	NCBI GENE
Annotations GO (BP)	NCBI GENE
Interactants	BIND, HPRD, via NCBI GENE
Outil de classement	GO-Family (GO ToolBox)
Données expérimentales (Analyses transcriptomiques, CGH, etc.)	Fichiers Excel « maison »

➔ *Mise en œuvre des connaissances du domaine, rôle de l'expert du domaine*

Modélisation ACGR (5)

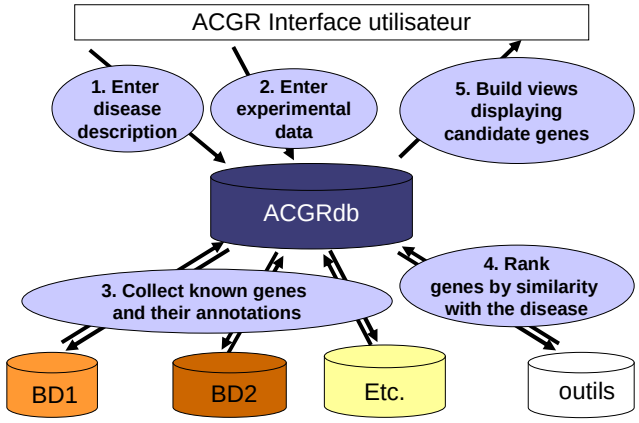
Modèle conceptuel de données



➔ *Mise en œuvre des connaissances du domaine, rôle de l'expert du domaine*

2. L'approche ACGR - Réalisation

Fonctionnalités mises en oeuvre



Création de la BD ACGR

Modèle relationnel de données

Gene (**Gene_ID**, Symbol, Organism, Complete_name, Chromosome, Cytoband, OMIM_ID,Source_ID)
GO_Term (**GO_ID**, Term, GO_section, Definition)
Gene_GO_Term (***Gene_ID, GO_ID***, Source_ID)
Orthology (***Gene_ID1, Gene_ID2***, Source_ID)
Interaction (***Gene_ID1, Gene_ID2***, Source_ID, Interaction_Type)
Disease (**Disease_ID**, Synopsis, OMIM_ID)
Disease_GO_Term (***Disease_ID, GO_ID,*** **Author_ID**)
Involvement (***Gene_ID, Disease_ID,*** Source_ID)
Ranking_Tool **(Tool_ID**, Description, URL)
Gene_Disease_Rank (***Gene_ID, Disease_ID, Author_ID, Tool_ID***, Value*)*
Experiment (**Exp_ID**, Type, Date, Platform, Analysis_procedure, *Disease_ID*)
Gene_Experiment (**Gene_ID, Exp_ID**, Ratio)

➔ ***Création de la base de données ACGR (MySQL)***

Peuplement de la BD ACGR
Wrappers

- Collecte de données dans les BD publiques
  - Système Xcollect
    - *http://www.loria.fr/~devignes/*
  - Définition de scénarios
    - *Pour chaque étape : URL de la ressource, requête, données à collecter délimitées dans le document HTML par des expressions régulières,*
  - Mapping des données collectées sur le modèle relationnel
- Outils de classement des gènes
  - Pour remplir la table Gene_Disease_Rank
  - Un wrapper par outil
  - Exemple du programme GO-Family de GO-ToolBox
    - *Pourcentage de termes communs entre les deux listes en incluant tous les ancêtres de chaque terme*

Création des vues ACGR (1)
Définitions

- Exemples de vues définies sur ACGRdb
  - Vue 1: Gènes classés selon leur similarité d'annotation GO avec la maladie
    - *Remarque : interclassement des gènes d'espèces différentes*
  - Vue 2: Orthologues humains des gènes d'organismes modèles classés selon leur similarité d'annotation GO avec la maladie
    - *Un gène candidat peut avoir un orthologue extrêmement bien annoté chez la souris ou la drosophile*
  - Vue 3: Gènes interactants des gènes de la Vue 1
  - Vue 4: Orthologues humains de gènes interactants des gènes des organismes modèles de la Vue 1
- Variantes avec données d'expression
  - Quand celles-ci sont disponibles

Création des vues ACGR (2)
Requêtes SQL

- **Vue 1 :** Gènes classés selon leur similarité d'annotation GO avec la maladie
  - **SQL expression (a est l'identifiant de la maladie concernée):**

```
CREATE VIEW View1 AS
   (SELECT g.Symbol, g.Organism, r.Value, g.Cytoband
   FROM Gene g, Gene-Disease-Rank r, Disease d
   WHERE g.Gene_ID = r.Gene_ID AND r.Disease_ID=d.disease_ID AND
   d.Disease_ID = a
   ORDER BY r.Value DESC)
```

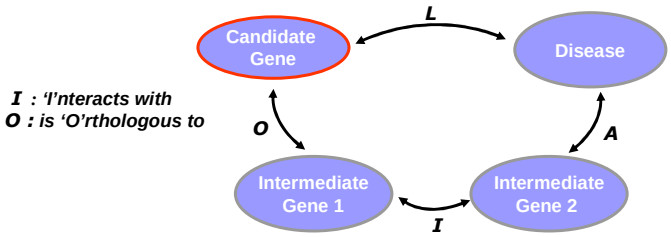
- **La vue est dynamique : exécutable après chaque mise à jour de la BD**
- **Autres vues disponibles sur http://bioinfo.loria.fr/projets/acgr**

3. Résultats pour 2 syndromes rares

- Syndrome de Goltz - FDH (OMIM #305600)
  - ***Hypoplasie dermique focale : atrophie et pigmentation anormale de la peau, hernies graisseuses à travers les défauts du derme, papillomes des muqueuses ou de la peau, anomalies des doigts.***
  - ***CGH-array : CNV en Xp11.23***
  - ***Gène responsable identifié en 2007 : PORCN (Porcupine)***
- Syndrome d'Aicardi (OMIM %304050)
  - ***Triade caractéristique : agénésie du corps calleux, anomalies chorio-rétiniennes, spasmes en flexion. Retard mental***
  - ***Génétique : chromosome X***
  - ***Gène responsable encore inconnu***
  - ***Etudié au Laboratoire de Génétique Humaine de la Faculté de Médecine de Nancy (Prof Jonveaux)***

Exemple de résultat ACGR (1)
Le syndrome de Goltz

- Description de la maladie
  - ***Skin defects*** → ***GO term « Skin development »***
  - ***Digit anomalies*** → ***GO term « Embryonic digit morphogenesis »***
  - ***Skeletal defects*** → ***GO term « Embryonic skeletal morphogenesis »***
- Modèle de gène candidat



I : 'I'nteracts with
O : is 'O'rthologous to

L : is co- 'L'ocalized on the genome
A : is 'A'nnotated in a similar way

Exemple de résultat ACGR (1)
Le syndrome de Goltz (suite)

- Tableau des premiers gènes de la BD ACGR-Goltz

Gène en interaction avec le gène orthologue au gène candidat | Gene Orthologue au gène candidat | Gène candidat

Symbol	Organism	Cytoband	Rank (%)	Interactant_Symbol	Source	Interactant_Cytoband	Interactant_Rank (%)	Human_Ortholog	Ortholog_Cytoband	Ortholog_Rank (%)
Wnt7a	mouse	6 39.5 cM	**31**	Porcn	BIND	X 2.15 cM	5	***PORCN***	**Xp11.23**	7
Ngfr	mouse	11 55.6 cM	22	Ndn	BIND	7 28.0 cM	11	*NDN*	15q11.2-q12	11
Ngfr	mouse	11 55.6 cM	22	Ndnl2	BIND	7 C	13	*NDNL2*	15q13.1	13

Exemple de résultat ACGR (2)
Le syndrome d'Aicardi

- **Input 1: disease description**
  - *Corpus callosum agenesis* → *5 GO terms « Forebrain development », « Corpus callosum development », « Corpus callosum morphogenesis », « neuron migration », « neural plate development ».*
  - *Chorio-retinal lacunae* → *1 GO term « Camera-type eye morphogenesis »*
  - *Infantile spasms* → *no appropriate GO term*
- **Input 2 : Dysregulated genes**
  - *ANOVA list of 300 genes (Saliha Yilmaz PhD Thesis)*
- **Input 3 : Copy Number Variation (CGH array)**

Exemple de résultat ACGR (2)
Le syndrome d'Aicardi (suite)

- **Vue 1 : Gènes classés par similarité avec la maladie et co-localisés ou dérégulés**

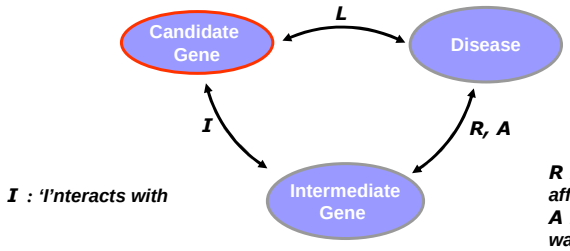
***L* : is co- 'L'ocalized on the genome**
***R* : is dys- 'R'egulated in affected patients**
***A* : is 'A'nnotated in a similar way**

Deux conclusions sur les listes de gènes obtenues dans la BD ACGR:

1. Les gènes sur le chromosome X qui sont bien classés ne sont pas dérégulés chez les patientes
2. Les gènes sur le chromosome X qui sont ré-régulés chez les patientes ne sont pas bien classés du point de vue de la similarité des annotations fonctionnelles

Exemple de résultat ACGR (2)
Le syndrome d'Aicardi (suite)

- **Vue 2 : Gènes qui interagissent avec un gène similaire ou dérégulé**



***I* : 'I'nteracts with**

***R* : is dys- 'R'egulated in affected patients**
***A* : is 'A'nnotated in a similar way**

- **Tableau des résultats de la Vue 2**

Gène en interaction avec le gène candidat — Gène candidat

Symbol	Species	CytoLoc	Rank (%)	Exp_ Value	Interactant	Source	Interactant_Cytoloc	Interactant_ Rank (%)	Interactant_ Exp_ Value
DLX5	human	7q22	50	1	*MAGED1*	HPRD	Xp11.23	3	0
UBE3A	human	15q11-q13	22	1	*UBQLN2*	HPRD	Xp11.23-p11.1	8	0
CXCL10	human	4q21	21	1	*CXCR3*	HPRD	Xq13	10	0
IGF1	human	12q22-q23	21	1	*IGSF1*	BIND	Xq25	6	0

Conclusion sur l'approche ACGR

- ACGR prototype
  - **Dans la BD ACGR, tout gène qui s'y trouve a une « bonne raison» d'y être.**
  - **ACGR permet d'intégrer des données expérimentales personnelles avec les données publiques**
  - **ACGR permet d'exploiter de nombreuses définitions de gènes candidats**
- Perspectives
  - **Séquencer les gènes candidats du syndrome d'Aicardi pour détecter des mutations (en cours)**
  - **Introduire d'autres méthodes de classement**
  - **Appliquer cette approche à d'autres types de gènes candidats : chémogénomique, médecine chinoise**

PLAN DE L'EXPOSÉ

I. Les différentes stratégies pour l'intégration de données

II. ACGR : Un entrepôt de données personnalisé pour l'identification de gènes candidats

III. MODIM : Généralisation à la construction de jeux de données pour la fouille de données
  I. **Principe: une aide à la construction d'entrepôts personnalisés**
  II. **Prototype: description**
  III. **Tests**

MODIM

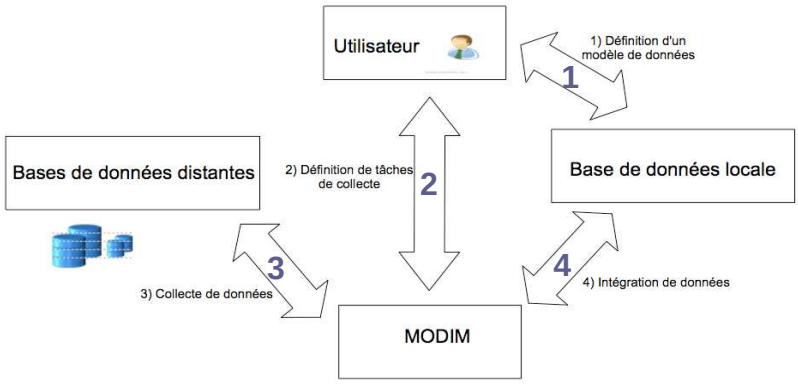
« Model-Driven Data Integration for Mining »

- Contexte de « Data Mining »: Fouille de données en vue de l'extraction de connaissances
  - **KDD (Knowledge Discovery in Databases) : 3 étapes itératives**

Objectifs de MODIM:

1. **Intégrer les données pertinentes pour le problème posé**
2. **Construire des jeux de données pour la fouille**

1. Aide à la construction d'entrepôts personnalisés (1)



1. Aide à la construction d'entrepôts personnalisés (2)

- Spécification du problème
  - **Par l'utilisateur, expert du domaine**
  - **Choix des données à collecter, des sources de données à interroger**
  - **Création d'un modèle conceptuel (~ représentation des connaissances)**
- Création de la BD MODIM
  - **Sauvegarde du modèle**
  - **S'appuie sur un système de gestion de BD**
- Configuration des tâches MODIM pour la collecte et la sauvegarde données
  - **Selon un modèle spécifié en XML**
- Exécution de ces tâches
  - **Peuplement de la BD**

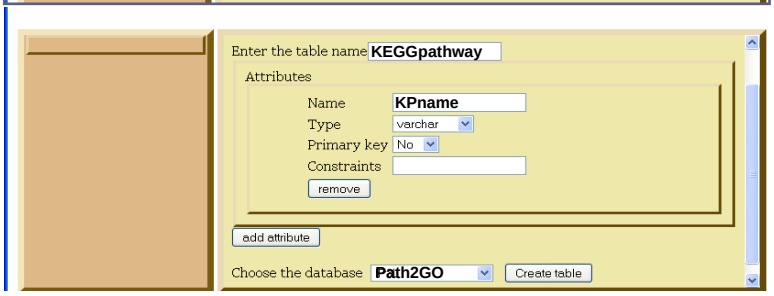
2. MODIM : Application web

Conçu et développé par Birama Ndiaye, IJD, Equipe Orpailleur INRIA Nancy Grand-Est

2.1 Création d'une BD

A

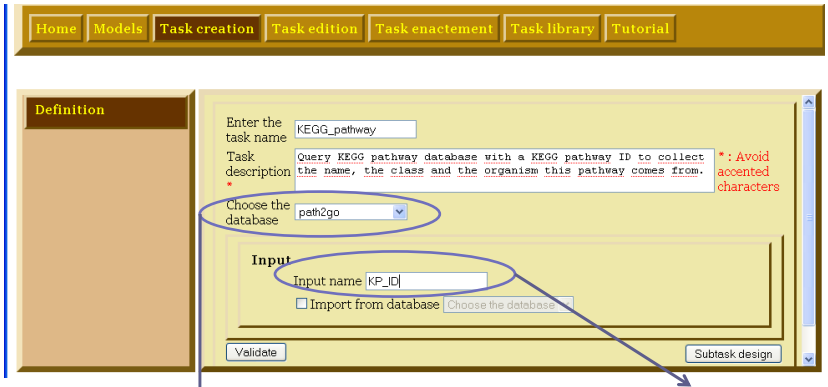
B



Etc.

2.2 Définition d'une tâche de collecte

Pour une BD et un type de donnée d'entrée

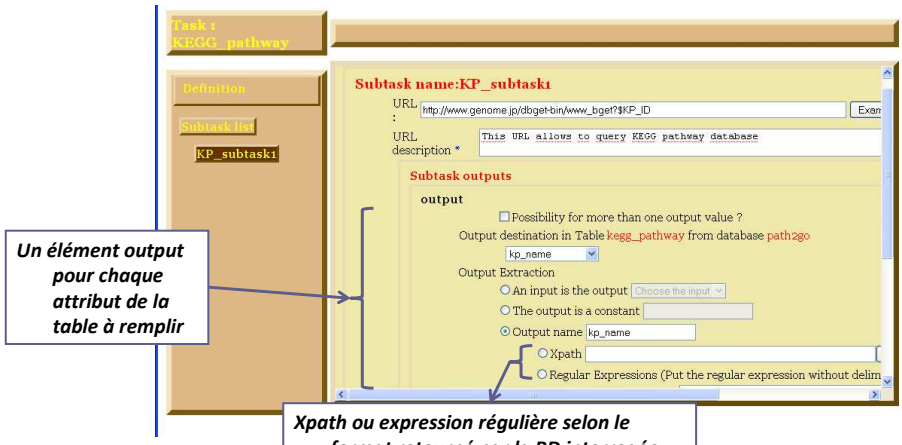


Spécifie la BD que l'on veut peupler

Spécifie le type d'entrée pour les requêtes associées à cette tâche

2.3 Définition d'une sous-tâche

Pour une BD distante à interroger



Un élément output pour chaque attribut de la table à remplir

Xpath ou expression régulière selon le format retourné par la BD interrogée

2.4 Mise en œuvre de la collecte

Moteur d'exécution des tâches et des sous-tâches programmé en PHP
Interrogation des BD spécifiées dans les sous-tâches
Extraction des données spécifiées
Ecriture dans la base de données concernée

MODIM : Bilan

- Aide à la réalisation de BD personnalisées à double titre
  - **Selon un modèle personnalisé**
  - **Avec possibilité d'importer des données personnelles**
- Accès à la BD via un système de gestion de BD relationnelle classique
  - **PostgreSQL**
  - **Interrogation en langage SQL**
  - **Possibilité de créer des « Vues » dynamiques comme avec ACGR**
  - **Possibilité d'enrichir l'entrepôt par des fonctionnalités complémentaires si nécessaire**
- Prototype en cours de test
  - **Stage de M1 : prise en main par une biologiste en 2 mois**
  - **Ouvert à collaboration si vous avez des données à collecter et à intégrer !**

CONCLUSION

- Proposition d'entrepôt personnalisé
  - **Approche générique et non généraliste**
  - **Garder l'utilisateur biologiste au cœur du système**
- Technologies du web sémantique
  - **Mise à disposition d'« entrepôts RDF » (ORACLE 10g & 11g)**
    - ***ex d'utilisation: AlzPharm***
  - **Données représentées par des graphes de triplets**
    - ***ex: (Gene ABC, HasSnp, rs23456)***
    - ***Langage de requête SPARQL***
- Rôle des ontologies
  - **Maîtriser leur utilisation à des points-clé du processus d'intégration**
  - **Systèmes à base de connaissance et de raisonnement : à venir…**
    - ***ne passent pas encore l'échelle en terme de quantité de données à traiter en biologie***

Participants

Laboratoire de Génétique
Faculté de Médecine Nancy

Saliha Yilmaz

Prof Philippe Jonveaux

Prof Bruno LeHeup

Thrombosis Research Institute
& King's College

John Louis McGregor and his team

LORIA, Equipe Orpailleur
Nancy

MD Devignes

Malika Smaïl-Tabbone

Birama Ndiaye (ingénieur CDD INRIA)

Cédric Bicep, Lu Zhang, Anaïs Gigant (stagiaires)

Financement

Communauté Urbaine du Grand Nancy

Contrat Plan Etat Région : Intelligence Logicielle

AAL association A.A.L. - Syndrome d'Aicardi

Lille, 4 mai 2010