

"Bioinformatique et ontologies"
Module MED, Montpellier, 22 mai 2008

Ontologies et Système d'information

Patrice Duroux

Ingénieur de Recherche CNRS
Laboratoire d'ImmunoGénétique Moléculaire
Université Montpellier, UPR CNRS 1142, IGH

- 1) Système d'information : une définition ?
- 2) Méthodologie et formalisme
- 3) IMGT® et système d'information
- 4) Avec outils et technologies
- 5) Étude de cas : IMGT/3Dstructure-DB
- 6) Conclusion

Systeme d'information

Informatique comme science du traitement de l'information

Besoin du concept de système

Systeme d'information \neq 1 logiciel ? N logiciels ?

Notion de ressources humaines et matérielles

Vu d'un ensemble organisé

Méthodologie et formalisme

Aujourd'hui : MERISE ou / et UML ?

MERISE : **M**éthode d'**E**tude et de **R**éalisation **I**nformatique
pour les **S**ystèmes d'**E**ntreprise

(fin années 70, fruit d'une consultation nationale du Ministère)

Méthodologie modèle conceptuel et physique

Approche données/traitements

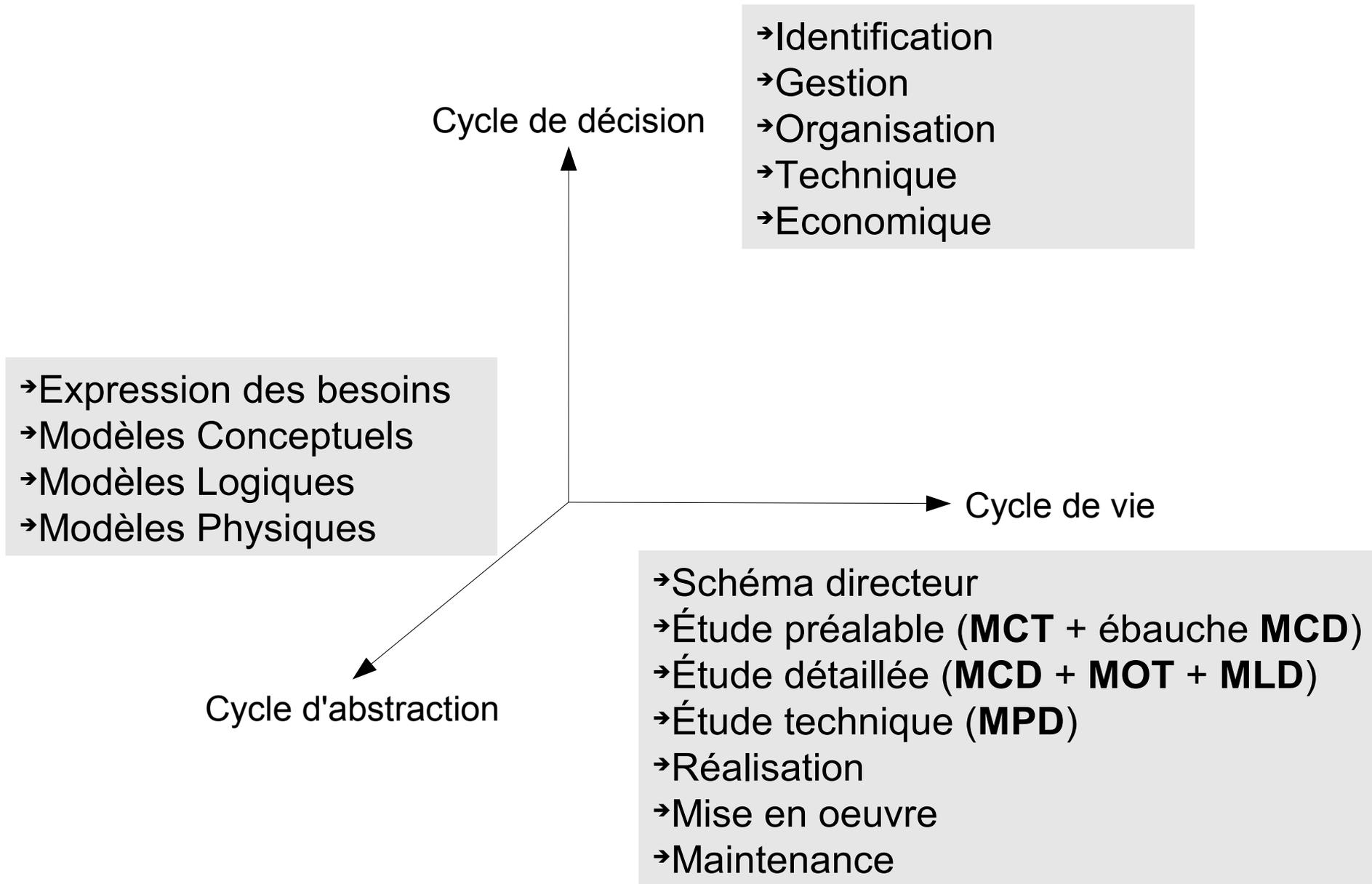
UML : **U**nified **M**odeling **L**anguage

(années 90, factorise des méthodes objets existantes : OMT,
Booch et OOSE)

Notation (type de diagrammes) et recommandation

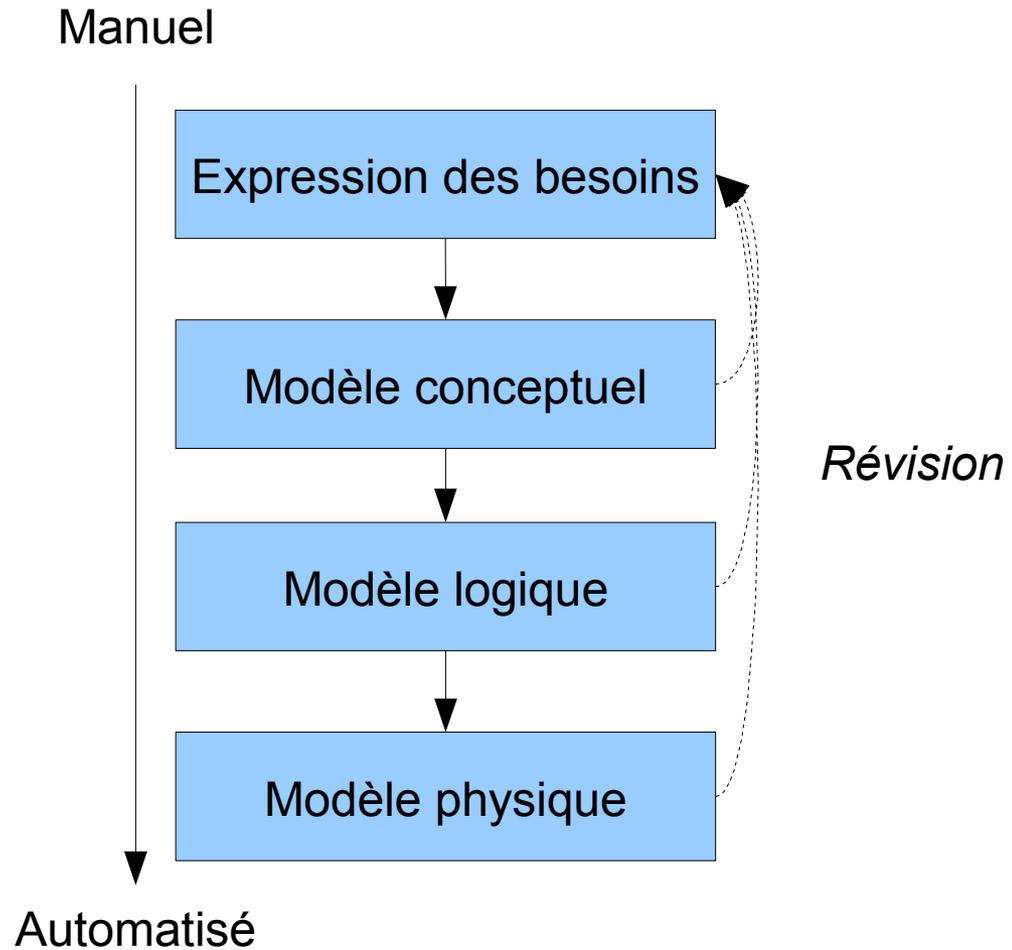
Approche objet

MERISE : démarche selon 3 axes



Cycle d'abstraction pour la conception

- Enchaînement de grandes étapes
- Vérifier la concordance des données et des traitements (oubli ? superflu ?)



IMGT® : quid ?

Une encyclopédie de connaissance et banque de données en immunogénétique ?

Un portail Internet ?

Une plate-forme ?

Un ensemble de services ?

Un système d'information ?

Ressource (1) : une équipe



Ressource (2) : un environnement informatique

Parc informatique : 3 serveurs pour le portail public, 3 serveurs et 15 postes pour le développement, le test et le portail Intranet.

Bases de données disponibles :

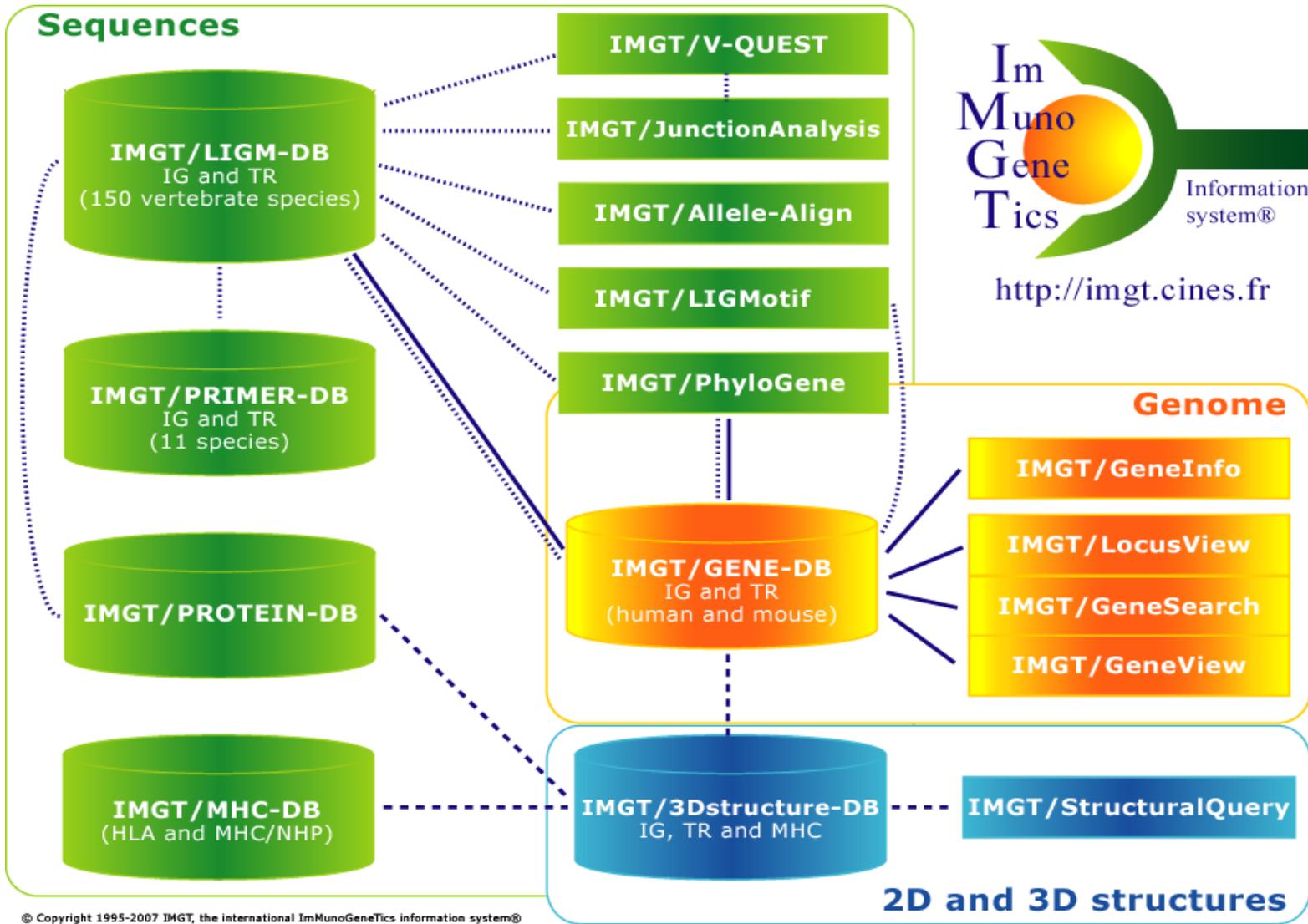
4 de séquences annotées : IMGT/LIGM-DB, IMGT/MHC-DB, IMGT/PRIMER-DB and IMGT/Protein-DB,

1 de gènes : IMGT/GENE-DB,

1 de structures 3-dimensionnelles : IMGT/3Dstructure-DB.

et 15 outils interactifs en ligne :

Bases de données et outils sur le portail

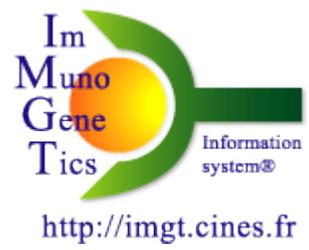


© Copyright 1995-2007 IMGT, the international ImMunoGeneTics information system®

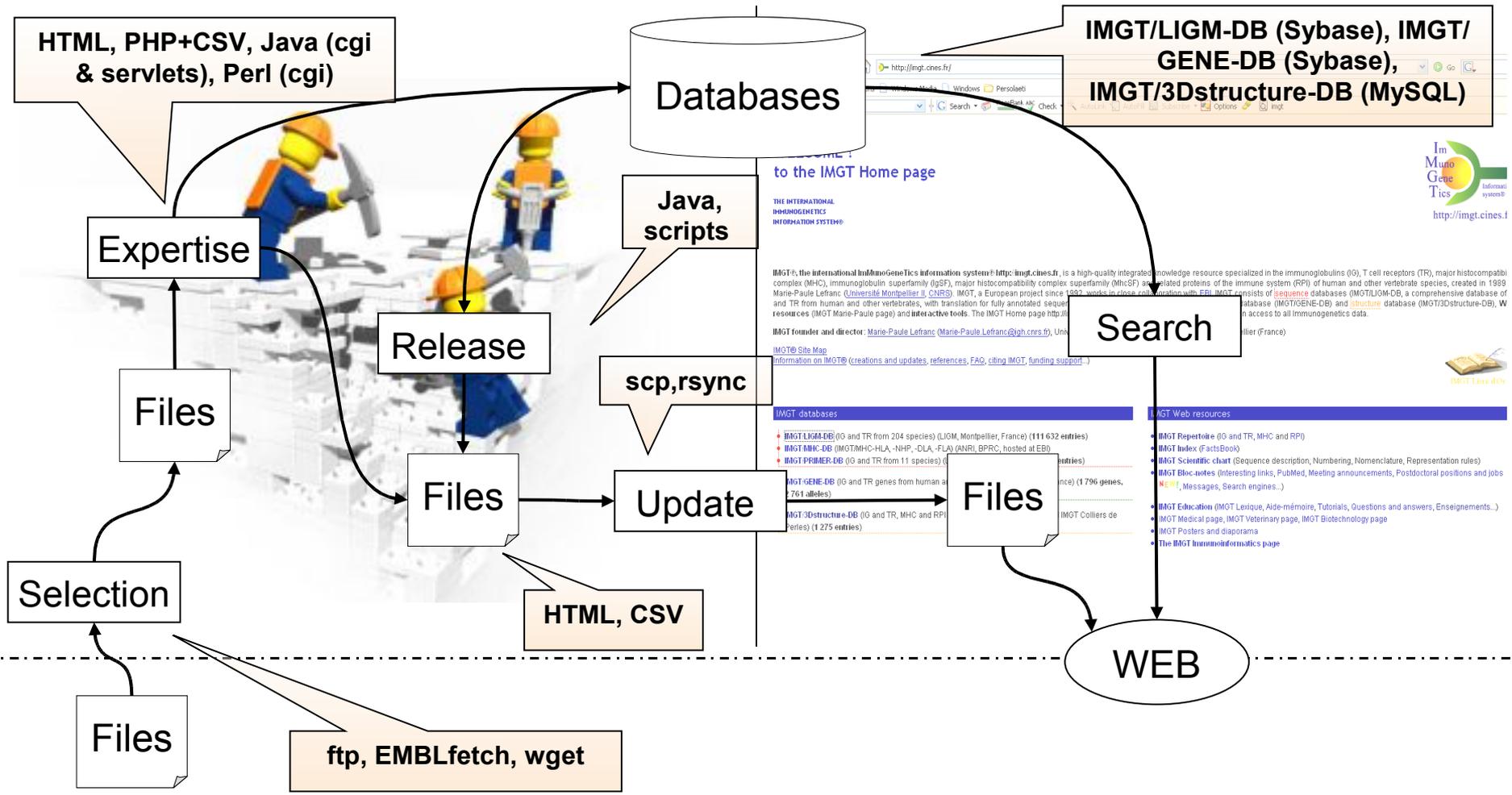
- Genetic approach
- Genomic approach
- - - - - Structural approach

MCT/MPD pour la mise à jour des données du portail

Production



Exploitation



LIGM-DB : MPD en chiffres

123 105 fiches EMBL, 1 base Sybase (3,9 Go)

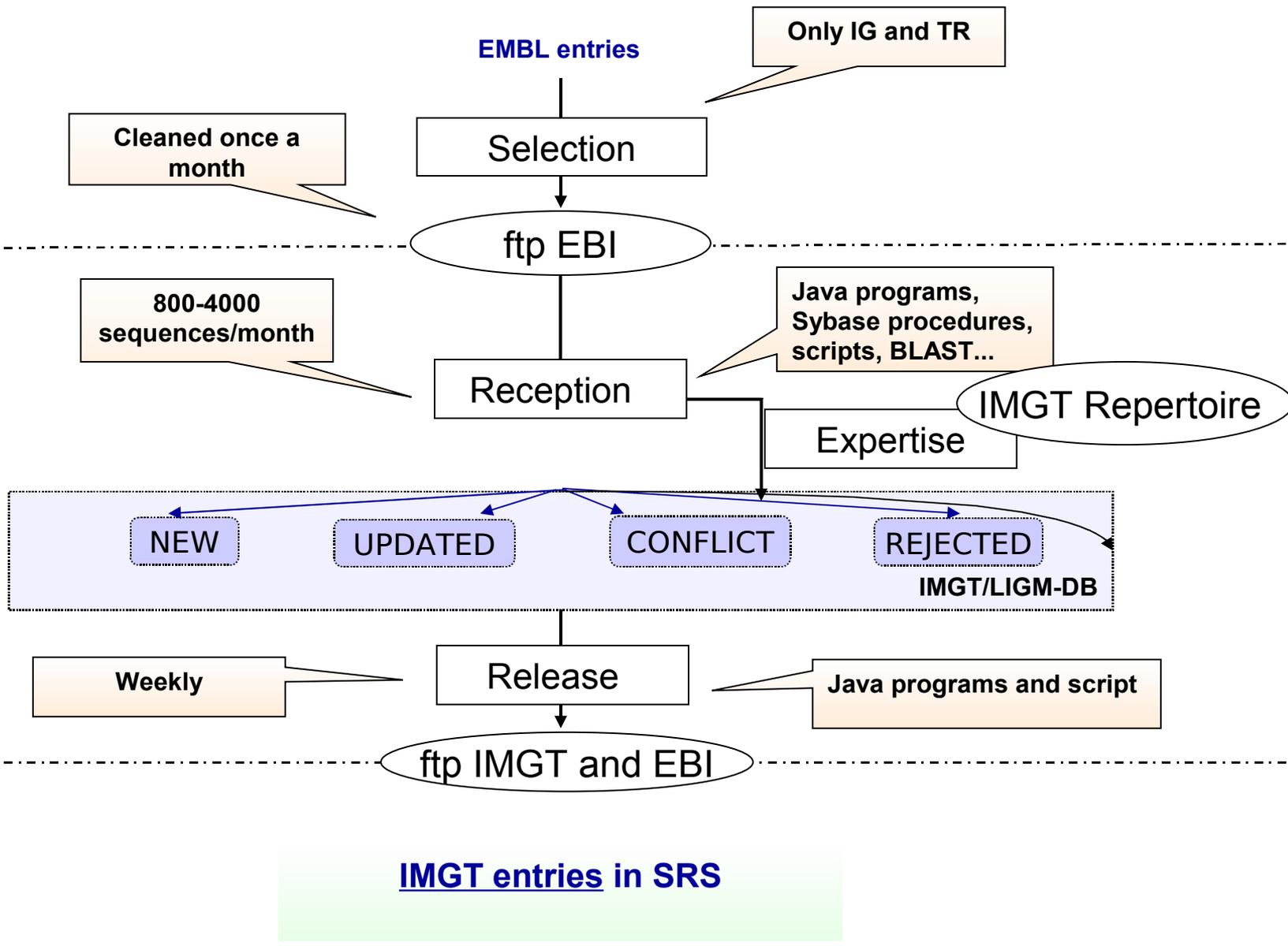
Attributs	230
Tables	105
Colonnes	527
Triggers	54
Procédures	21
Espèces	231
Séquences	123 105
K	95 558
A	27 547
Keywords	184
Labels	232
Specificities	670
Features	509 038
Nucleotides	102 397 151

IMG/IMG-DB (with EMBL)

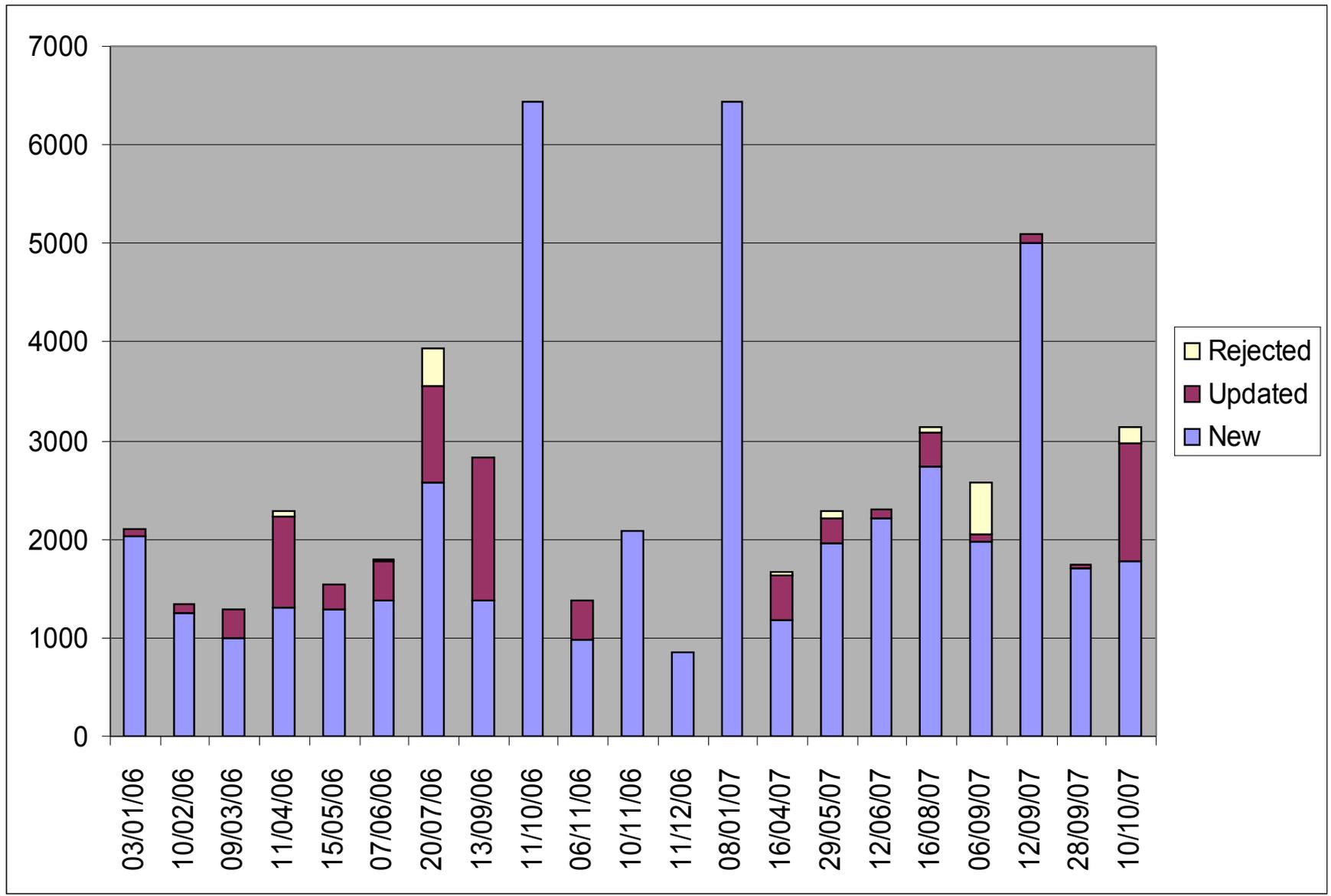
EMBL

IMG

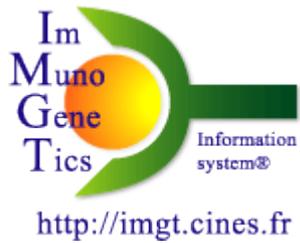
EMBL



IMGT/LIGM-DB : mouvement des données



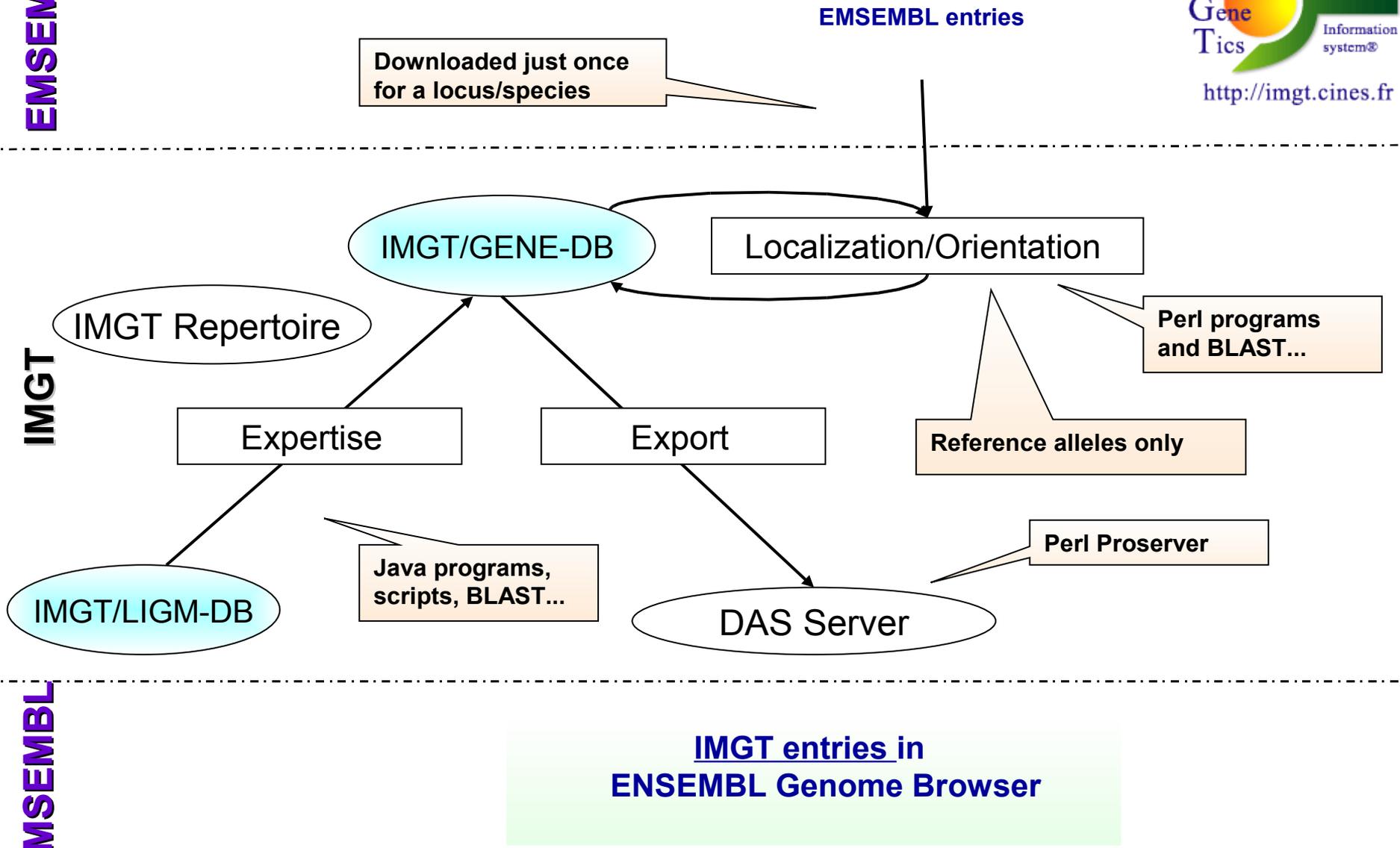
IMG/GENE-DB (with ENSEMBL)



EMSEMBL

IMG

EMSEMBL



GENE-DB : MPD en chiffres

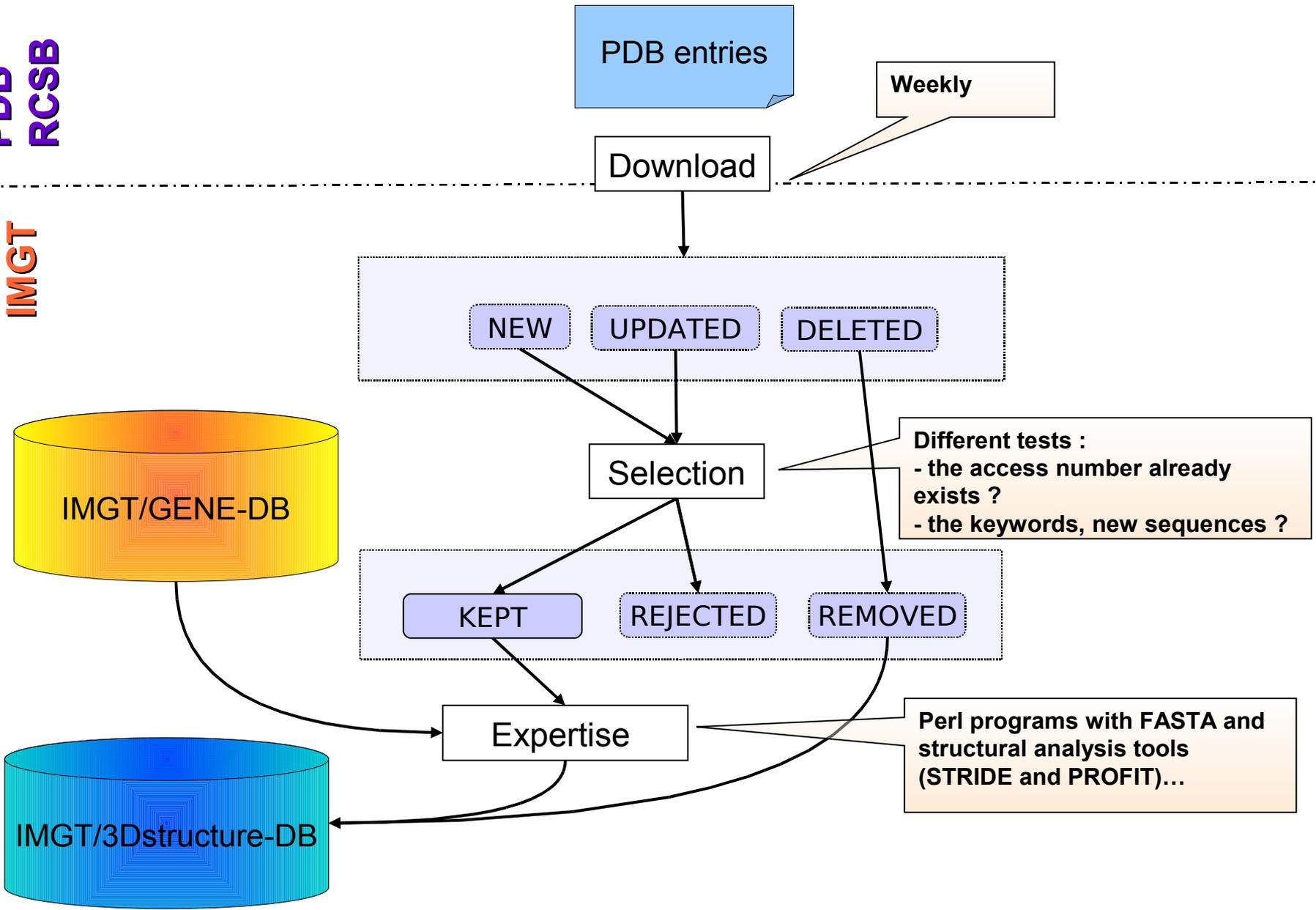
1 base Sybase (52 Mo)

Attributs	105
Tables	51
Colonnes	422
Triggers	45
Procédures	6
Espèces	8
Gènes	1 900
Allèles	2 894
Nucléotides	729 768

IMG/3Dstructure-DB (with PDB)

PDB
RCSB

IMG

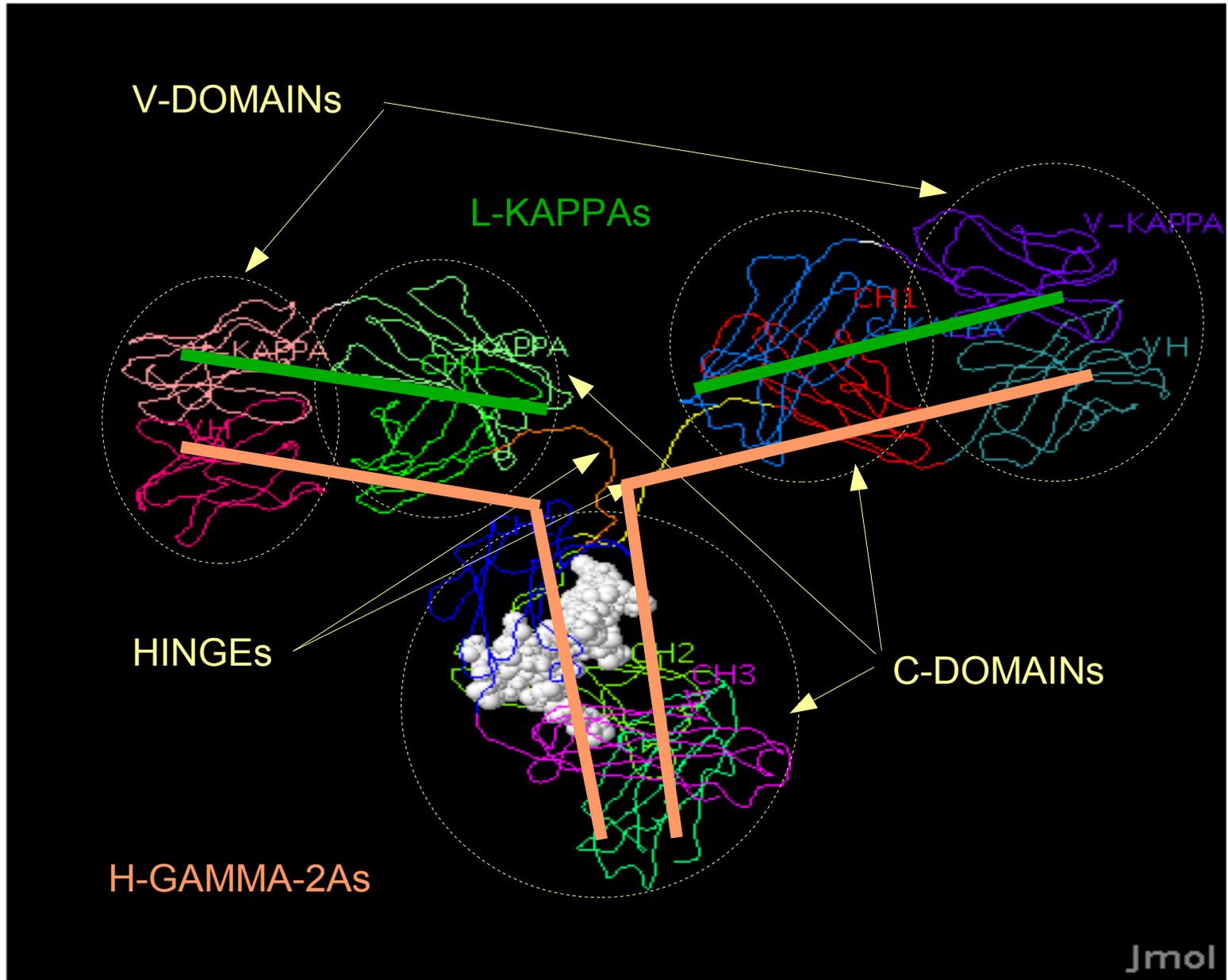


LIGM-DB : MPD en chiffres

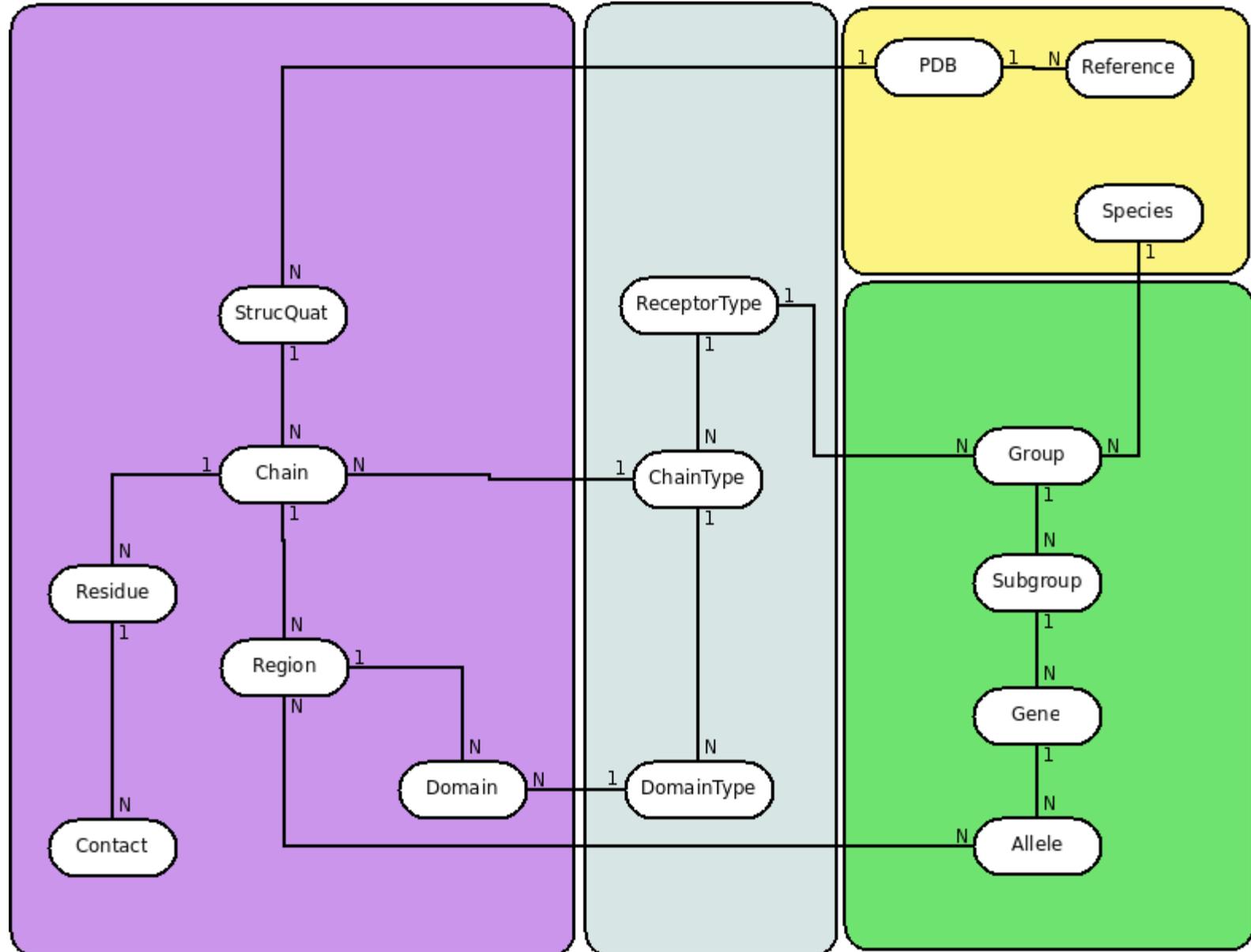
123 105 fiches EMBL, 1 base Sybase (3,9 Go)

Attributs	230
Tables	105
Colonnes	527
Triggers	54
Procédures	21
Espèces	231
Séquences	123 105
K	95 558
A	27 547
Keywords	184
Labels	232
Specificities	670
Features	509 038
Nucleotides	102 397 151

Exemple : la structure 1IGT



3Dstructure-DB : un pré MCD



3Dstructure-DB : un fragment du MPD

ReceptorType(receptortypeid,receptortype)

MainReceptorDescr(mainreceptordescrid,definition,label)

ReceptorDescr(receptordescrid,@mainreceptordescrid,definition,label)

MainDomDescr(maindomdescrid,definition,label)

DomDescr(domdescrid,@maindomdescrid,definition,label,domtype)

ChainDescr(chaindescrid,definition,label)

Dom_ChainDescr(domchaindescrid,@domdescrid,@chaindescrid,numdom)

Chain_ReceptorDescr(chainreceptordescrid,@chaindescrid,@receptordescrid,numchain)

Groupe(groupid,groupe,@receptortypeid)

Espece(especeid,latin,english,abb)

Subgroup(subgroupid,subgroup,@groupid,@especeid)

Gene(geneid,gene,@groupid,@subgroupid,@especeid)

Conclusion

Systeme d'information pour IMGT : un long cheminement

Outil conceptuel de collaboration

Opportun à la Démarche Qualité et la gestion de projets

Ontologie : un bon support pour les modèles de données